

VŠB – Technická univerzita Ostrava  
Fakulta elektrotechniky a informatiky  
Katedra kybernetiky a biomedicínského inženýrství

**Využití adaptivního Neuro-Fuzzy  
inferenčního systému ve zpracování  
signálů v oblasti zpracování řeči**

**Using the Adaptive Neuro-Fuzzy  
Inferential System in Signal Processing  
in Speech Processing**

# Zadání diplomové práce

Student: **Bc. Adam Holub**

Studijní program: N2649 Elektrotechnika

Studijní obor: 2612T041 Řídicí a informační systémy

Téma: **Využití adaptivního Neuro-Fuzzy inferenčního systému ve zpracování  
signálů v oblasti zpracování řeči**  
**Using the Adaptive Neuro-Fuzzy Inferential System  
in Signal Processing in Speech Processing**

Jazyk vypracování: čeština

Zásady pro vypracování:

Cílem diplomové práce je využití neuro-fuzzy inferenčního systému v oblasti zpracování řeči a implementace řešení do programovacího jazyka C#.

Body zadání:

1. Rozbor funkce neuro-fuzzy inferenčního systému a pochopení jeho funkce.
2. Rozbor problematiky zpracování řeči.
3. Implementace neuro-fuzzy inferenčního systému s využitím stávajících implementací v jazyce C#.
4. Provedení experimentálních testů v oblasti zpracování řeči.
5. Statistické zhodnocení provedených testů.
6. Zhodnocení závěrečné práce.

Seznam doporučené odborné literatury:

- [1] NAGEL, Christian, Jay GLYNN a Morgan SKINNER. *Professional C# 5.0 and .NET 4.5.1*. Indianapolis, IN: John Wiley and Sons, 2014. ISBN 978-1-118-83294-3.
- [2] VLACHÝ, Václav. *Praxe zvukové techniky*. 2. aktualiz. vyd. Praha: Muzikus, c2000. ISBN 80-86253-05-8.
- [3] JANG, Jyh-Shing R. ANFIS: adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics*. May/Jun 1993, vol.23, no.3, pp.665-685. ISSN 0018-9472. doi: 10.1109/21.256541.

Formální náležitosti a rozsah diplomové práce stanoví pokyny pro vypracování zveřejněné na webových stránkách fakulty.

Vedoucí diplomové práce: **Ing. Jaromír Konečný, Ph.D.**

Konzultant diplomové práce: doc. Ing. Radek Martinek, Ph.D.

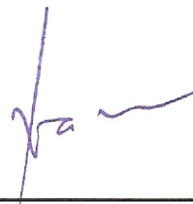
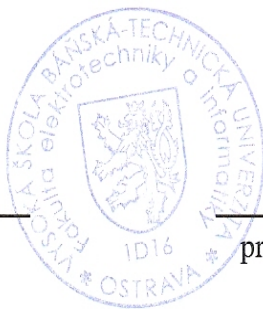
Datum zadání: 01.09.2017

Datum odevzdání: 30.04.2018



---

doc. Ing. Jiří Koziorek, Ph.D.  
*vedoucí katedry*



---

prof. Ing. Pavel Brandštetter, CSc.  
*děkan fakulty*

Prohlašuji, že jsem tuto diplomovou práci vypracoval samostatně. Uvedl jsem všechny literární  
prameny a publikace, ze kterých jsem čerpal.

V Ostravě 30. dubna 2018

Adm. Hruš  
.....

Rád bych poděkoval svému vedoucímu Ing. Jaromírovi Konečnému, Ph.D. za odborné vedení a vstřícnost při psaní této diplomové práce. Dále bych rád poděkoval doc. Ing. Radkovi Martinkovi, Ph.D. za cenné rady při konzultacích.

## **Abstrakt**

Tato diplomová práce se zabývá návrhem softwarového řešení adaptivního systému ANFIS s cílem jeho využití v oblasti zpracování zašuměné řečové nahrávky. V práci je popsán návrh adaptivního systému a jeho implementace v programovacím jazyce C#. Efektivita vytvořené aplikace je otestována na syntetických i reálných datech se zaměřením na kvalitu filtrace pro různé parametry ANFIS. Pro objektivní posouzení efektivity filtrace jsou stanoveny hodnotící kritéria SNR, SSNR, DTW a RMSE.

**Klíčová slova:** ANFIS, řeč, šum, adaptivní systém, filtrace, LMS, RLS, SNR, SSNR, DTW, RMSE, neuronové sítě, C#

## **Abstract**

This master's thesis deals with a design of a software application of adaptive system ANFIS with aim of use in a noise suppression of speech record. Design and implementation of this adaptive system is described in this work. Efficiency of created application is tested on synthetic and real data with the focus on filtration efficiency under different ANFIS parameters. For objective evaluation of filtration efficiency SNR, SSNR, DTW and RMSE are estimated.

**Key Words:** ANFIS, speech, noise, adaptive system, filtering, LMS, RLS, SNR, SSNR, DTW, RMSE, neural network, C#

# Obsah

<b>Seznam použitých zkratk a symbolů</b>	<b>9</b>
<b>Seznam obrázků</b>	<b>11</b>
<b>Seznam tabulek</b>	<b>13</b>
<b>Úvod</b>	<b>14</b>
<b>1 Zvuk a zpracování řeči</b>	<b>15</b>
1.1 Deformace zvuku v prostředí . . . . .	16
1.1.1 Odstup signálu od šumu . . . . .	17
1.2 Zpracování řeči . . . . .	18
1.2.1 Jednokanálové metody . . . . .	19
1.2.2 Vícekanálové metody . . . . .	20
<b>2 Fuzzy logika</b>	<b>21</b>
2.1 Funkce příslušnosti . . . . .	21
2.2 IF-THEN pravidla . . . . .	22
2.3 Systém fuzzy inference . . . . .	22
<b>3 Metody adaptivní filtrace</b>	<b>24</b>
3.1 LMS a RLS algoritmus . . . . .	24
3.2 Umělé neuronové sítě . . . . .	26
<b>4 ANFIS</b>	<b>29</b>
4.1 Architektura . . . . .	30
4.2 Učení . . . . .	32
4.2.1 Dopředný chod . . . . .	32
4.2.2 Zpětný chod . . . . .	33
<b>5 Využití ANFIS ve zpracování řeči</b>	<b>35</b>
5.1 Potlačení šumu v nahrávce řeči . . . . .	35
5.2 Porovnání s metodami HHT a WT . . . . .	36
<b>6 Implementace systému ANFIS</b>	<b>37</b>
6.1 Architektura a návrh systému . . . . .	37
6.1.1 Prezentační vrstva . . . . .	39
6.1.2 Aplikační vrstva . . . . .	40

<b>7</b>	<b>Experimenty</b>	<b>41</b>
7.1	Metody hodnocení kvality filtrace . . . . .	41
7.1.1	SNR a segmentální SNR . . . . .	41
7.1.2	RMSE . . . . .	43
7.1.3	DTW . . . . .	43
7.2	Testování na umělých datech . . . . .	43
7.3	Filtrace řeči . . . . .	45
7.3.1	Deformace šumu FIR filtrem . . . . .	46
7.3.2	Deformace šumu nelineární funkcí . . . . .	50
7.3.3	Reálná data . . . . .	54
<b>8</b>	<b>Zhodnocení výsledků</b>	<b>60</b>
	<b>Závěr</b>	<b>62</b>
	<b>Literatura</b>	<b>63</b>
	<b>Přílohy</b>	<b>65</b>
<b>A</b>	<b>Obsah přiloženého CD</b>	<b>66</b>



## Seznam použitých zkratk a symbolů

ADALINE	– Adaptivní lineární neuron ( <b>A</b> daptive <b>L</b> inear <b>N</b> euron)
AM	– Amplitudová Modulace
ANFIS	– Adaptivní Neuro-Fuzzy Inferenční Systém ( <b>A</b> daptive <b>N</b> euro- <b>F</b> uzzy <b>I</b> nference <b>S</b> ystem)
BP/GD	– Zpětné šíření/Sestupný gradient ( <b>B</b> ack <b>P</b> ropagation/ <b>G</b> radient <b>D</b> escent)
DP	– Dolní propust
DTW	– <b>D</b> ynamic <b>T</b> ime <b>W</b> arping
FIR	– Konečná impulsní odezva ( <b>F</b> inite <b>I</b> mpulse <b>R</b> esponse)
FIS	– Fuzzy Inferenční Systém ( <b>F</b> uzzy <b>I</b> nference <b>S</b> ystem)
FM	– Frekvenční Modulace
GUI	– Grafické uživatelské rozhraní ( <b>G</b> raphical <b>U</b> ser <b>I</b> nterface)
HHT	– Hilbert-Huang Transformace ( <b>H</b> ilbert- <b>H</b> uang <b>T</b> ransform)
HP	– Horní propust
ICA	– Analýza nezávislých komponent ( <b>I</b> ndependent <b>C</b> omponent <b>A</b> nalysis)
IIR	– Nekonečná impulsní odezva ( <b>I</b> nfinite <b>I</b> mpulse <b>R</b> esponse)
$L_I$	– Hladina intenzity zvuku
LMS	– Metoda nejmenších čtverců ( <b>L</b> east <b>M</b> edian <b>S</b> quare)
$L_P$	– Hladina akustického tlaku
MF	– Funkce příslušnosti ( <b>M</b> embership <b>F</b> unction)
MSE	– Střední kvadratická odchylka ( <b>M</b> ean <b>S</b> quare <b>E</b> rror)
N	– Třetí vrstva ANFIS
NN	– Neuronová síť ( <b>N</b> eural <b>N</b> etwork)
PCM	– Pulsně kódová modulace ( <b>P</b> ulse <b>C</b> ode <b>M</b> odulation)
PP	– Pásmová propust
PZ	– Pásmová zadrž
RLS	– <b>R</b> ecursive <b>L</b> east <b>S</b> quare
RMS	– Efektivní hodnota ( <b>R</b> oot <b>M</b> ean <b>S</b> quare)
RMSE	– Relativní střední kvadratická odchylka ( <b>R</b> oot <b>M</b> ean <b>S</b> quare <b>E</b> rror)
SNR	– Odstup signálu od šumu ( <b>S</b> ignal-to- <b>N</b> oise <b>R</b> atio)
SSNR	– Segmentální odstup signálu od šumu ( <b>S</b> egmental <b>S</b> ignal-to- <b>N</b> oise <b>R</b> atio)
VAD	– Detekce řečové aktivity ( <b>V</b> oice <b>A</b> ctivity <b>D</b> etection)
WAV	– <b>W</b> aveform Audio File Format
WT	– Vlnková transformace ( <b>W</b> avelet <b>T</b> ransform)

X	–	Univerzum
dB	–	decibel
$\Pi$	–	Druhá vrstva ANFIS
$\Sigma$	–	Pátá vrstva ANFIS

## Seznam obrázků

1	Šíření zvuku podélné (vlevo) a příčné (vpravo) [15] . . . . .	15
2	Odrazy zvuku v přirozeném prostředí (zjednodušeno). Upraveno z [11] . . . . .	17
3	Dozvuk v přirozeném prostředí [11] . . . . .	17
4	Přenosový kanál [19] . . . . .	18
5	Trojúhelníková (vlevo) a trapezodiální (vpravo) funkce [13] . . . . .	21
6	Gaussovká (vlevo) a sigmoidální (vpravo) funkce [13] . . . . .	22
7	Schéma fuzzy inferenčního systému [1] . . . . .	23
8	Obecné blokové schéma adaptivního systému. Upraveno z [20] . . . . .	24
9	Blokové schéma adaptivního filtru [28] . . . . .	25
10	Základní model neuronu [25] . . . . .	26
11	Typy aktivačních funkcí . . . . .	27
12	Schéma vícevrstvé NN [24] . . . . .	27
13	Mechanismus rozhodování dle modelu Takagi-Sugeno. Upraveno z [1] . . . . .	30
14	Struktura ANFIS. Upraveno z [20] . . . . .	30
15	Koeficienty definující tvar zvonové funkce. Upraveno z [1] . . . . .	34
16	Schéma adaptivního systému . . . . .	37
17	Diagram aktivit . . . . .	38
18	Uživatelské prostředí aplikace . . . . .	39
19	VAD v aplikaci . . . . .	42
20	Dynamic Time Warping . . . . .	43
21	Čistý sinusový signál (nahore), bílý šum (uprostřed) a zašuměný signál (dole) . .	44
22	Vyfiltrovaný sinusový signál . . . . .	45
23	Čistá řeč . . . . .	45
24	Reálný šum (nahore) a zašuměný signál (dole) . . . . .	46
25	Závislost na počtu fuzzy pravidel - deformace FIR filtrem . . . . .	47
26	Závislost SSNR na počtu epoch a počtu fuzzy pravidel - deformace FIR filtrem .	48
27	Závislost DTW na počtu epoch a počtu fuzzy pravidel - deformace FIR filtrem .	49
28	Vyfiltrovaná řeč - deformace FIR filtrem . . . . .	49
29	Bílý šum (nahore) a zašuměný signál (dole) . . . . .	50
30	Závislost na počtu epoch - deformace nelineární funkcí . . . . .	51
31	Závislost na počtu fuzzy pravidel - deformace nelineární funkcí . . . . .	51
32	Závislost SSNR na počtu epoch a počtu fuzzy pravidel - deformace nelineární funkcí	53
33	Závislost DTW na počtu epoch a počtu fuzzy pravidel - deformace nelineární funkcí	53
34	Vyfiltrovaná řeč - deformace nelineární funkcí . . . . .	54
35	Reálné nahrávky . . . . .	55
36	Závislost na počtu epoch - reálná data . . . . .	56
37	Závislost na počtu fuzzy pravidel - reálná data . . . . .	56

38	Závislost SSNR na počtu epoch a počtu fuzzy pravidel - reálná data . . . . .	58
39	Závislost DTW na počtu epoch a počtu fuzzy pravidel - reálná data . . . . .	58
40	Vyfiltrovaná řeč - Gaussova a zvonová MF . . . . .	59

## Seznam tabulek

1	Úrovně hluku v jednotkách dB a Pa. Upraveno z [11] . . . . .	16
2	Zpracování řeči v aplikacích [16] . . . . .	19
3	Vstupní charakteristiky signálu deformované FIR filtrem . . . . .	46
4	Výsledky pro jednotlivé kombinace počtu epoch a pravidel (zvonová MF) - deformace FIR filtrem . . . . .	47
6	Výstupní charakteristiky signálu deformované FIR filtrem . . . . .	49
7	Vstupní charakteristiky signálu pro vytvořený bílý šum . . . . .	50
8	Výsledky pro jednotlivé kombinace počtu epoch a pravidel (zvonová MF) - deformace nelineární funkcí . . . . .	52
10	Výstupní charakteristiky signálů pro vytvořený bílý šum . . . . .	54
11	Vstupní charakteristiky reálných nahrávek . . . . .	55
12	Výsledky pro jednotlivé kombinace počtu epoch a pravidel (Gaussova MF) - reálná data . . . . .	57
14	Výstupní charakteristiky signálů pro reálná data . . . . .	59
15	Čas zpracování . . . . .	61

## Úvod

Komunikace přes mobilní telefon je každodenní součástí našeho života. Procházející signál je však často degradován šumem na pozadí, kterým může být hluk na rušných společenských akcích, hluk v továrních halách nebo zvuky běžícího motoru při cestě autem. V takových případech je nutné zavést odpovídající filtr, který je schopen šum ze záznamu odfiltrovat a tím jej zkvalitnit. Problémem, který se v této situaci vyskytuje je nelineární charakter šumu, který určuje samotné prostředí, ve kterém šum vzniká. V přirozeném prostředí vznikají zvukové odrazy a dozvuky, které se následně přičítají k původnímu šumu. Je tedy zřejmé, že původní (zdrojový) šum nelze pouze odečíst.

V současné době je v oblasti potlačení šumu nejvíce využívaným řešením aplikace adaptivních algoritmů LMS a RLS. Algoritmus LMS je matematicky jednoduchý a výpočetně nenáročný, avšak při filtraci dosahuje větší odchylky od hledaného signálu. Druhý jmenovaný, algoritmus RLS, dosahuje oproti LMS lepších výsledků za cenu větší matematické náročnosti.

V reálných podmínkách algoritmy LMS a RLS trpí jedním nedostatkem, a tím je jejich lineární pojetí. V přirozeném prostředí často čelíme nelineárnímu charakteru šumu (deformace zvuku o překážky), který nemůže být zcela potlačen pouze aplikací lineárního filtru. Tato diplomová práce je zaměřena na využití struktury ANFIS (adaptivní neuro-fuzzy inferenční systém) pro potlačení šumu v zašuměné nahrávce řeči. ANFIS představuje kombinaci fuzzy inferenčního systému a umělých neuronových sítí, který je schopen modelovat hledaný nelineární systém a předvídat tak vývoj v čase proměnných signálů.

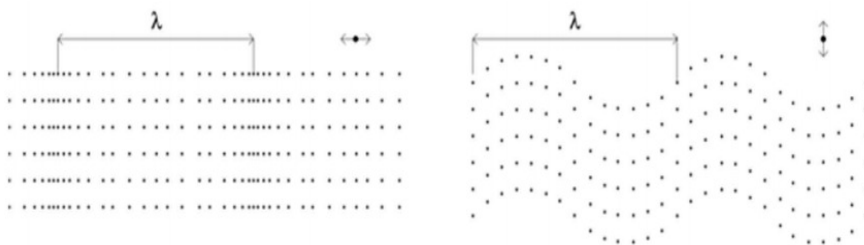
Cílem této diplomové práce je rozbor a pochopení funkce ANFIS a jeho aplikace v oblasti odšumení nahrávky řeči. Pro otestování systému budou provedeny experimenty zaměřené na filtraci signálů s rozdílně nastavenými parametry ANFIS, a také na hledání optimálních parametrů. Implementace systému ANFIS proběhne v programovacím jazyku C#.

Textová část této práce je rozčleněna na osm částí. První kapitola této práce se bude zabývat na fyzikální pohled na zvuk, jeho charakteristiky a deformace v prostředí. Dále bude rozebrána problematika zpracování řeči a náhled na algoritmy, které slouží pro zkvalitnění řečového signálu. Druhá kapitola poskytne popis fuzzy logiky, která je nedílnou součástí ANFIS. Vysvětleny budou pojmy jako funkce příslušnosti, podmíněná pravidla a fuzzy inference. V následující kapitole budou rozebrány techniky adaptivní filtrace. Budou zde popsány adaptivní algoritmy LMS a RLS, a dále principy neuronové sítě, která společně s fuzzy logikou tvoří základ ANFIS. Čtvrtá kapitola je věnována samotnému systému ANFIS. Systém bude analyzován z hlediska jeho architektury a učení. V páté kapitole budou zmíněny dosavadní studie, které se zabývaly experimentálním zpracováním zašuměného řečového signálu pomocí ANFIS. Šestá kapitola se již bude věnovat praktické části této práce. Bude zde uveden popis implementace systému ANFIS v programovacím jazyku C#. Na šestou kapitolu naváže kapitola sedmá, která bude zaměřena na zpracování zvukových nahrávek ve vytvořené aplikaci. Poslední kapitola bude věnována zhodnocení provedených experimentů a dosažených výsledků.

# 1 Zvuk a zpracování řeči

Zvuk je z fyzikálního hlediska mechanické vlnění, které se šíří pevným, plynným i kapalným látkovým prostředím. Frekvence vlnění, které je člověk schopen vnímat jako sluchový vjem se pohybuje v rozmezí 16 Hz až 20 kHz. Dolní mez je však u každého člověka individuální a může se pohybovat v rozmezí 16 až 20 Hz. Zvuky, které leží mimo toto slyšitelné pásmo, sluchové vjemy nevyvolává. Podle umístění ve frekvenčním spektru tyto zvuky rozdělujeme na infrazvuk a ultrazvuk. Infrazvuk má frekvenci nižší než je dolní hranice slyšitelného pásma (16 až 20 Hz), dolní hranicí je frekvence blízká nule. Naopak ultrazvuk dosahuje frekvencí vyšších než je horní hranice slyšitelného pásma, kterou je 20 kHz [9].

Zvuk se šíří rozkmitáváním částic daného prostředí okolo rovnovážné polohy. Kmitáním se předává energie sousedním částicím, čímž dochází k šíření kmitu. V kapalných a plynných prostředích se kmitání šíří podélně, v pevných prostředích lze pozorovat i šíření příčné [9].



Obrázek 1: Šíření zvuku podélné (vlevo) a příčné (vpravo) [15]

$\lambda$  označuje vlnovou délku, která určuje vzdálenost mezi dvěma částicemi se stejnou fází. Vlnová délka  $\lambda$  se vypočítá z podílu rychlosti šíření  $v$  a frekvencí kmitání  $f$ : [9, 10]

$$\lambda = \frac{v}{f} \quad (1.1)$$

Zvuk lze charakterizovat pomocí akustických veličin. Mezi základní veličiny patří frekvence, barva (složení) a hlasitost (intenzita). Jednotkou hlasitosti je bel [B], pojmenovaná po americkém vynálezci Alexandrovi Grahamovi Bellovi. Sama o sobě je tato jednotka velká, proto se v praxi hlasitost zvuku častěji vyjadřuje pomocí o řád nižších decibelů [dB]. Při šíření zvuku prostředím dochází k změnám tlaku vzduchu střídající oblasti akustického podtlaku a přetlaku. Tyto změny vnímáme subjektivně jako zvuk o určité hlasitosti. Měřitelnou veličinou akustického tlaku je hladina akustického tlaku  $L_p$  [dB], která je dána vztahem [10]:

$$L_p = 20 \cdot \log \frac{p}{p_0} \quad (1.2)$$

kde  $p$  označuje sledovanou efektivní hodnotu akustického tlaku a  $p_0$  tzv. práh slyšitelnosti (pro vzduch  $p_0 = 2 \cdot 10^{-5} \text{ Pa}$ ). Pro objektivní hodnocení zvuku používáme hladinu intenzity zvuku  $L_I$  [dB]. Vztah pro výpočet je obdobný jako u hladiny akustického tlaku, tentokrát se však vychází z intenzity zvuku, která je přímo úměrná energii kmitání. Ve vztahu  $I$  vyjadřuje

sledovanou hodnotu intenzity zvuku a  $I_0$  referenční hodnotu, která odpovídá prahu slyšitelnosti (pro vzduch  $I_0 = 10^{-12} \text{ W} \cdot \text{m}^{-2}$ ). [10, 11]

$$L_I = 10 \cdot \log \frac{I}{I_0} \quad (1.3)$$

Pro obě charakteristiky je poměr akustických tlaků resp. intenzit velmi malý, proto je vhodné vyjádřit je pomocí logaritmické stupnice.

Lidské ucho je schopno vnímat zvuk až do úrovně 130 - 140 dB, přičemž prahem bolesti pro některé může být již hranice 100 dB. Jednotlivé hladiny hluku a jejich hodnoty v jednotkách dB a Pa shrnuje tab. 1.

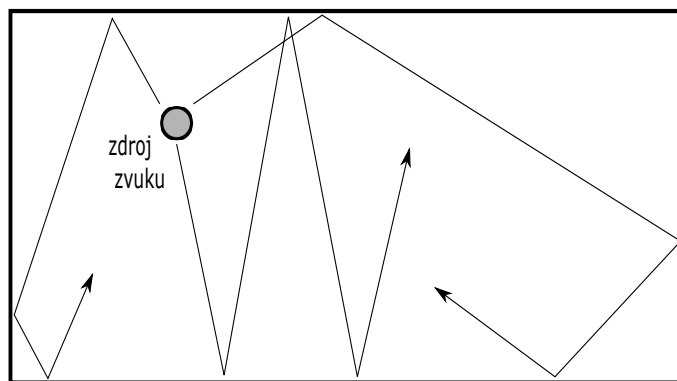
Tabulka 1: Úrovně hluku v jednotkách dB a Pa. Upraveno z [11]

	akustický tlak (dB)	akustický tlak (Pa)
práh slyšitelnosti	0	$2 \cdot 10^{-5}$
šum v izolované studiu	20	$2 \cdot 10^{-4}$
tíkot hodin	30	$6,3 \cdot 10^{-4}$
tichá ulice	40	$2 \cdot 10^{-3}$
šept ve vzdálenosti 10 cm	50	$6,3 \cdot 10^{-3}$
akustická kytara ve vzdálenosti 40 cm	60	$2 \cdot 10^{-2}$
saxofon	90	$6,3 \cdot 10^{-1}$
hlasitý zpěv (15 cm)	100	2
hlasitý výkřik přímo před ústy (práh bolestivosti)	130	63
vzlet tryskáče	190	20000

## 1.1 Deformace zvuku v prostředí

Zvuk se v prostředí šíří od zdroje zvuku všemi směry (kulovitě) rychlostí přibližně 340 m/s. Platí, že s rostoucí vzdáleností slábne intenzita zvuku. Dvojnásobku vzdálenosti od zdroje odpovídá útlum 6 dB, pro vzdálenost 4 m od zdroje je útlum 12 dB, pro vzdálenost 32 m je to 30 dB atd. Při šíření zvukové vlny dochází k její deformaci při nárazu na nějakou překážku, v takovém případě dochází částečně k pohlcení akustické energie překážkou a přeměnou na teplo, a částečně k odrazům a lámání zvuku. V přirozeném prostředí dochází k mnohonásobným odrazům, čímž dochází k interferenci originálního (původního) zvuku s těmito odraženými zvuky. Dochází tak k prodloužení a zesílení hladiny intenzity zvuku, tento jev nazýváme přirozeným dozvukem [11].





Obrázek 2: Odrazy zvuku v přirozeném prostředí (zjednodušeno). Upraveno z [11]

Odražené zvuky k nám přicházejí s různým zpožděním a z různých směrů. Při působení zdroje zvuku v prostoru je tento prostor vyplněn akustickou energií, která je z části pohlcena materiály, které se nacházejí v daném prostředí. Od okamžiku vypnutí zdroje zvuku bude zdrojem zvuku zmíněná energie, která však rychle klesá. Čas, za kterou hladina intenzity zvuku v místnosti klesne o 60 dB se nazývá dobou dozvuku [9, 11].



Obrázek 3: Dozvuk v přirozeném prostředí [11]

### 1.1.1 Odstup signálu od šumu

Při šíření signálu v přirozeném prostředí dochází k interferenci užitečného signálu s šumem. Šum představuje parazitní signály v prostředí, které svým působením znečišťují přenášený signál a tím zkreslují užitečnou informaci. Šum lze z hlediska jeho charakteru rozdělit na stacionární a nestacionární. Stacionární šum má spektrální výkonovou charakteristiku konstatní v čase (zvuk počítače, hluk klimatizace), zatímco nestacionární představuje náhodné zvuky a má tuto charakteristiku proměnnou (bouchnutí, projíždějící dopravní prostředek) [19].

Pro objektivní posouzení toho, jak je moc signál zašuměn lze stanovit parametr SNR (ang. Signal-to-noise ratio) neboli odstup signálu od šumu. Tento parametr vyjadřuje, kolikrát je výkon signálu větší než výkon šumu. Jedná se o bezrozměrnou veličinu definovanou vztahem: [12]

$$\text{SNR} = \frac{P_{\text{signal}}}{P_{\text{noise}}} \quad (1.4)$$

V praxi však signály většinou mají velmi široký rozsah a bývají vyjadřovány v logaritmických jednotkách dB. Pro tyto účely lze vztah přepsat do tvaru vyjadřující hodnotu SNR v decibelech [12, 17]:

$$\text{SNR}_{\text{dB}} = 10 \cdot \log\left(\frac{P_{\text{signal}}}{P_{\text{noise}}}\right) \quad (1.5)$$

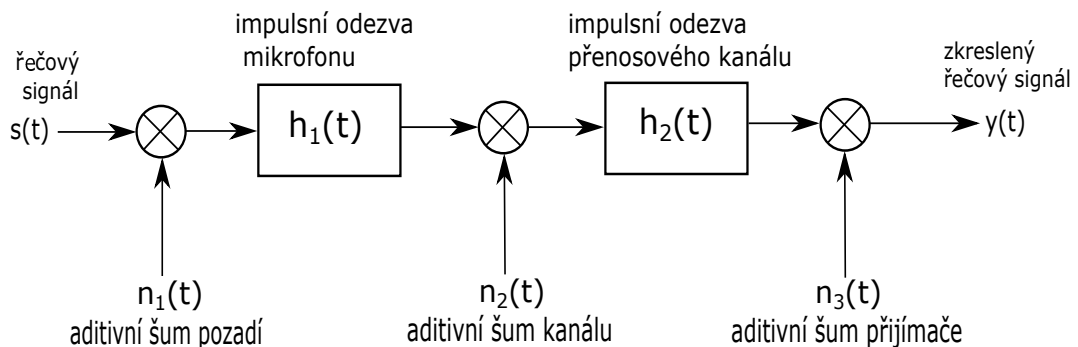
Z uvedeného vztahu vyplývá, že čím vyšší bude hodnota SNR, tím vyšší bude odstup signálu od šumu a tím kvalitnější bude signál. Platí, že hodnota  $\text{SNR} = 3 \text{ dB}$  (poměr 2:1) bývá označována jako hranice mezi nízkým a vysokým SNR [12].

## 1.2 Zpracování řeči

Řeč je akustický nestacionární signál nacházející se ve slyšitelném pásmu zvukového spektra (16 Hz - 20 kHz) a skládá se především z akustické složky a lingvistické informace [19]. Na stupnici intenzity dosahuje lidská řeč hodnot v rozmezí 40 až 60 dB, přičemž křik může nabývat hodnot až kolem 80 dB [9, 10].

Řečový signál je v komunikačních systémech přenášen skrze přenosový kanál. Kvalita přenosu řeči je dána kvalitou přenosového kanálu. Mezi nejdůležitější součásti patří měnič, který akustický signál snímá a převádí na signál elektrický. Nejčastěji používaným měničem je mikrofon, který se však nikdy nehodí pro všechny účely, a je tedy nutné pro danou situaci zvolit vhodný typ na základě jeho vlastností (citlivost, směrová a frekvenční charakteristika,...). Volba správného mikrofonu výrazně ovlivňuje výslednou kvalitu řečového signálu [11].

Přenos signálu kanálem bývá z pravidla narušen aditivním šumem, který je k signálu přičítán po celé délce přenosového kanálu [19]. Schéma přenosového kanálu je uvedeno na obr. 4



Obrázek 4: Přenosový kanál [19]

V telefonních a rozhlasových aplikacích je důležitým faktorem, který ovlivňuje kvalitu přenášené informace, šířka frekvenčního pásma. V telefonních aplikacích je běžně používána hodnota 3,2 kHz (300 Hz - 3,5 kHz), zatímco u AM a FM rádiových vysíláních, kde je požadován kvalitní přenos řeči, je šířka pásma nastavena na 8 kHz. Vzorkovací frekvence je 8 kHz pro telefonní aplikace a 16 kHz pro rádiové vysílání. Při 8 bitech na vzorek je přenosová rychlost 64 Kbps ( $8000 \text{ vzorků} \cdot 8 \text{ bit}$ ) resp. 128 Kbps [16]. Tyto hodnoty přehledně shrnuje následující tabulka:

Tabulka 2: Zpracování řeči v aplikacích [16]

Aplikace	Šířka pásma (Hz)	Vzorkovací frekvence (Hz)	Přenosová rychlost (Kbps)
Telefonní přenos	3200	8000	64
Rádiové vysílání	8000	16000	128

Předtím než je řeč zpracována (např. v aplikacích pro rozpoznání řeči) jsou na ni aplikovány algoritmy pro její zkvalitnění. Tyto algoritmy slouží pro zvýšení srozumitelnosti řeči a potlačení obsaženého šumu s co nejmenší ztrátou užitečné informace. Algoritmy pro zkvalitnění řečového signálu lze rozdělit do dvou základních kategorií: jednokanálové a vícekanálové metody.

Jednokanálové metody vyžadují pouze jeden vstup v podobě zašuměného signálu (snímání pouze jedním mikrofonom). Tyto metody potlačují šum na základě rozdílných charakteristik užitečného signálu a šumu. Druhou kategorií jsou multikanálové metody. Snímání probíhá pomocí dvou a více mikrofonů, kde jeden, primární, snímá zašuměný signál a ostatní, referenční, snímají pouze šum [31, 32].

### 1.2.1 Jednokanálové metody

Jednokanálové metody nacházejí své uplatnění v situacích, kde druhý snímací kanál není k dispozici. Tyto systémy jsou v porovnání s vícekanálovými jednodušší i levnější. Metody jsou založené na předpokladu, že užitečný signál (řeč) a šum v pozadí mají rozdílné charakteristiky. Ve svých implementacích využívají výpočtu frekvenčního spektra signálu z dílčích segmentů signálu. Efektivita těchto metod je obvykle omezena při přítomnosti nestacionárního šumu [26, 32].

- frekvenčně selektivní filtry typu FIR a IIR (DP, HP, PP, PZ),
- metody založené na teorii Wienerovy filtrace,
- spektrální odčítání,
- Kalmanova filtrace.

Metoda spektrálního odčítání patří mezi nejjednodušší metody potlačení šumu. Spektrální odčítání je založeno na stanovení amplitudového nebo výkonového spektra obou signálů (zašuměné řeči a šumu), které jsou následně od sebe odečteny. Spektrum šumu je určeno z úseků neobsahujících řeč.

Ačkoliv metoda spektrálního odčítání dokáže šum z řeči potlačit, trpí jistými nedostatky. Je-li totiž metoda vázána na amplitudové resp. výkonové spektrum, spektrum fázové zůstane stejné i po filtraci. Dále efektivita metody je velmi závislá na určení spektra šumu, což je v reálných podmínkách obtížný úkol [31, 33].

Se stanovením spekter pracuje také Wienerův filtr. Ten patří do kategorie filtrů, které minimalizují střední kvadratickou chybu (MSE) mezi zašuměným a čistým řečovým signálem. Spektrum hledaného signálu je stanoveno podle následující rovnice [33]:

$$|\hat{X}(k)| = H(k)|Y(k)|, \quad (1.6)$$

kde  $H(k)$  představuje  $k$ -tou komponentu Wienerova filtru a  $Y(k)$  je spektrum zašuměného signálu.  $Y(k)$  je dáno součtem spekter obou signálů, což vychází z předpokladu, že šum je aditivní složkou řeči. Filtr  $H(k)$  můžeme stanovit následovně [33]:

$$H(k) = 1 - \frac{|N(k)|^2}{|Y(k)|^2} \quad (1.7)$$

Alternativní přístupem stanovení Wienerova filtru je využití a priori SNR.

$$H(k) = \frac{\xi_k}{1 + \xi_k}, \quad (1.8)$$

kde  $\xi_k$  je a priori SNR  $k$ -té komponenty frekvenčních spekter signálů. Tento přístup rozdělí signál na úseky s vysokou hodnotou SNR, které jsou zachovány, a úseky s malou hodnotou SNR, které jsou potlačeny [33].

### 1.2.2 Vícekanálové metody

Systémy využívající multikanálové metody těží z přítomnosti jednoho či více referenčních mikrofónů, které snímají pouze šum. Tyto systémy jsou v porovnání s jednokanálovými systémy více komplexnější, avšak dosahují kvalitnějších výsledků, zejména při přítomnosti nestacionárního šumu. Do této skupiny patří [31, 32]:

- metoda analýzy nezávislých komponent (ICA),
- adaptivní metody (LMS, RLS, umělé neuronové sítě, ANFIS),
- metoda tvarování paprsku.

Skupina adaptivních metod patří mezi metody s velmi dobrou filtrační schopností využívající adaptivního algoritmu ve svém jádru. Tento algoritmus je implementován za účelem úpravy parametrů filtru tak, aby se filtr přizpůsoboval aktuálním podmínkám. Vybrané adaptivní metody budou blíže popsány v kapitole 3.

## 2 Fuzzy logika

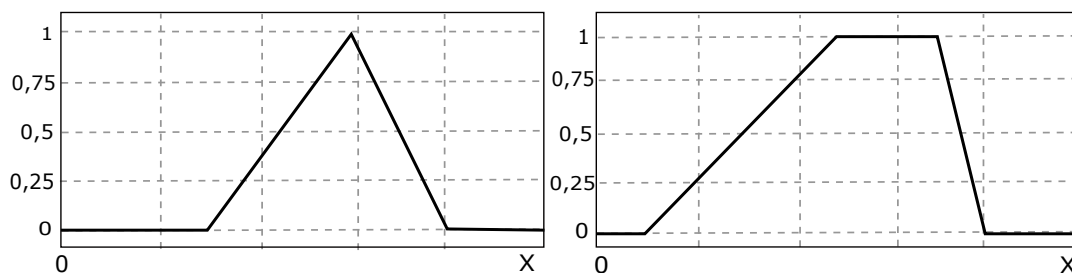
Základy fuzzy logiky definoval již v roce 1965 profesor L.A. Zadeh. Ve svém článku [27] představuje novou kategorii logiky, která umožňuje rozlišit pojmy pravda/nepravda jinak, než bylo dosud zvykem v klasické teorii. V klasické teorii množin jsou uvažovány pouze dvě hodnoty, 1 a 0. Tyto hodnoty nám určují s jistotou zda prvek do dané množiny patří (logická 1) nebo nepatří (logická 0), nepřipouští tedy žádnou nejasnost, neurčitost nebo kompromisy. Fuzzy logika tyto omezení nemá a pracuje s celým intervalem  $\langle 0,1 \rangle$ . Díky fuzzy logice můžeme tedy prohlásit, že prvek do určité množiny patří (a zároveň nepatří) jen částečně. Tuto částečnost definuje tzv. míra příslušnosti prvku dané třídě. Soubor všech prvků, jejichž míra příslušnosti není 0 nebo 1, ale leží v tomto intervalu nazýváme fuzzy setem [13, 14].

$$A = (X, \mu_A) \quad (2.1)$$

Fuzzy množina je definována dvojicí uvedenou v rovnici 2.1.  $X$  se nazývá univerzem<sup>1</sup> množiny  $A$  a  $\mu_A$  je funkcí příslušnosti.

### 2.1 Funkce příslušnosti

Hodnota z intervalu  $\langle 0,1 \rangle$  je každému prvku přiřazena na základě funkce příslušnosti. Jedná se o křivku definovanou na celém rozsahu od 0 do 1. V asociaci s fuzzy setem funkce příslušnosti mapuje vstupní hodnotu na odpovídající míru příslušnosti (tzv. fuzzifikace vstupních ostrých hodnot). Typů funkcí příslušnosti je mnoho, nejjednodušší z nich je trojúhelníková nebo trapezodiální, které tvoří pouze rovné čáry [13].



Obrázek 5: Trojúhelníková (vlevo) a trapezodiální (vpravo) funkce [13]

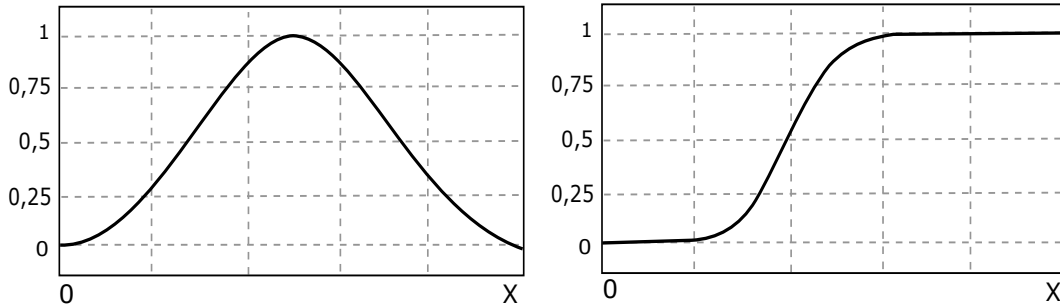
Příklad matematického popisu trojúhelníkové funkce [26]:

$$\mu_A(x) = \begin{cases} 0 & x < \alpha \\ (x - \alpha)/(\beta - \alpha) & \alpha \leq x \leq \beta \\ (\gamma - x)/(\gamma - \beta) & \beta \leq x \leq \gamma \\ 0 & x > \gamma \end{cases}, \quad (2.2)$$

<sup>1</sup>Všechny možné hodnoty vstupních (ostrých) proměnných do fuzzy systému

kde  $x$  představuje ostrou hodnotu,  $\alpha$  bod počátku vzestupné hrany,  $\beta$  vrchol trojúhelníkové funkce a  $\gamma$  bod konce sestupné hrany.

Další typy funkcí příslušnosti jsou již matematicky složitější s hladším průběhem. Mezi ně patří Gaussovská funkce, sigmoidální nebo zvonová funkce [13].



Obrázek 6: Gaussovská (vlevo) a sigmoidální (vpravo) funkce [13]

Příklad matematického popisu Gaussovské MF [1]:

$$\mu_A(x) = \exp\left(-\frac{(c-x)^2}{2\sigma^2}\right), \quad (2.3)$$

kde  $x$  představuje ostrou hodnotu,  $c$  vrchol funkce, a  $\sigma$  šířku funkce.

## 2.2 IF-THEN pravidla

Pro formální zápis lidské znalosti, resp. popisu chování systému, lze použít jazykový model. Ten je definován jako složitý fuzzy výrok obsahující jazykové proměnné, jména jejich hodnot, logické spojky a hodnoty výroků z intervalu  $\langle 0,1 \rangle$  [14]. Důležitou roli hrají tzv. podmíněná tvrzení formulující IF-THEN pravidla, které mají tvar [13]:

$$\text{IF } x \text{ is } A, \text{ THEN } y \text{ is } B, \quad (2.4)$$

kde  $x$  a  $y$  představují jazykové proměnné,  $A$  a  $B$  jsou jazykovými hodnotami na intervalu  $\langle 0,1 \rangle$ . Část výrazu vlevo od THEN se nazývá předpokladem neboli premisou, část za slovem THEN je důsledkem neboli konsekventem [14].

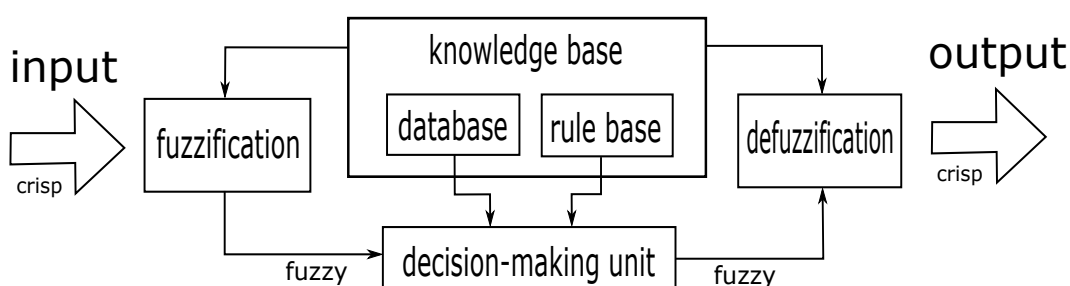
Pomocí IF-THEN pravidel dochází k fuzzifikaci vstupních hodnot a jeho výstupem je fuzzy set, který je reprezentován funkcí příslušnosti. Počet výrazů v premise není omezen, pro případy dvou a více výrazů jsou aplikovány logické operátory (AND, OR, NOT) [13, 14].

## 2.3 Systém fuzzy inference

Fuzzy inferenční systém je proces mapování vstupních hodnot na hodnoty výstupní využívající fuzzy logiku. Poskytuje základnu pro rozhodovací proces a rozpoznávání vzorů. Existují dva

základní typy systému, Mamdani-type a Sugeno-type, liší se ve způsobu určování výstupu. Mamdani-type očekává na výstupu fuzzy set, který je následně defuzzifikován. Systém typu Sugeno modeluje na výstupu funkce příslušnosti, které jsou lineární funkcí nebo konstantou (není potřeba dělat defuzzifikaci) [13, 14].

Systém fuzzy inference se skládá z pěti bloků: blok fuzzifikace, který transformuje ostrá data na fuzzy množiny, blok dat, který definuje funkce příslušnosti, blok báze pravidel, obsahující definovaná pravidla, rozhodovací jednotku, provádějící inferenci dat s pravidly a blok defuzzifikace, který převádí výsledný fuzzy set na výstupní ostrou hodnotu. Jak bylo zmíněno výše, poslední blok defuzzifikace se nachází pouze u FIS typu Mamdani. Výsledná ostrá hodnota může být určena např. jako průměr maxim nebo těžiště plochy výstupní funkce příslušnosti [1, 14].

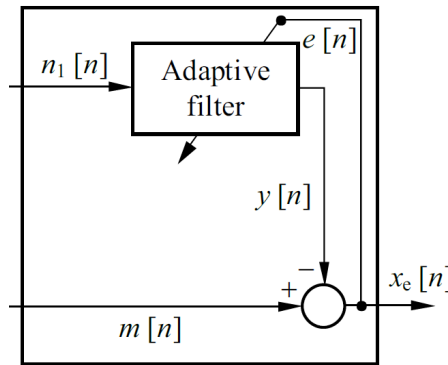


Obrázek 7: Schéma fuzzy inferenčního systému [1]

### 3 Metody adaptivní filtrace

Reálné fyzikální aplikace jsou zatíženy spoustou nelinearit a neurčitostí. Ve většině případů požadujeme, aby filtr pracoval v neznámém a často časově proměnném prostředí, jehož vývoj se dá těžko odhadovat. Pro odstranění těchto měnících se podmínek je nutné použít filtr, který je schopen se těmto změnám efektivně přizpůsobovat. Filtry, které takovou vlastnost mají nazýváme adaptivními filtry.

Základní vlastnost adaptivního filtru spočívá v kontinuální změně parametrů podle předem stanoveného algoritmu s využitím informace ze vstupu i pomocné informace, přicházející většinou z výstupu celého systému. V systému, který využívá adaptivní filtr, je základním prvkem zpětná vazba, pomocí které dochází v úpravám parametrů samotného filtru. Cílem je snížení účelové funkce na její minimum. Tato funkce je závislá pouze na hodnotě odchylky  $e$  [17, 21].



Obrázek 8: Obecné blokové schéma adaptivního systému. Upraveno z [20]

Na obr. 8 je uvedeno obecné schéma systému s adaptivním filtrem, kde  $x_e(n)$  představuje výstupní signál systému,  $y(n)$  výstupní signál filtru,  $m(n)$  měřený zašuměný signál,  $n(n)$  šum z referenčního snímače a  $e(n)$  reprezentuje odchylku výstupního od měřeného signálu.

Metody adaptivní filtrace lze rozdělit do dvou základních skupin: lineární filtry (LMS, RLS, Kalmanův filtr, ADALINE) a nelineární filtry (Vícevrstvé neuronové sítě, Evoluční algoritmy) [18]. Lineární filtrace je odvozena od lineárního časově invariantního systému, neboli systému, ve kterém platí princip superpozice. Nelineární systémy jsou takové, ve kterých tento princip neplatí. V následujících kapitolách budou stručně popsány vybrané metody adaptivní filtrace.

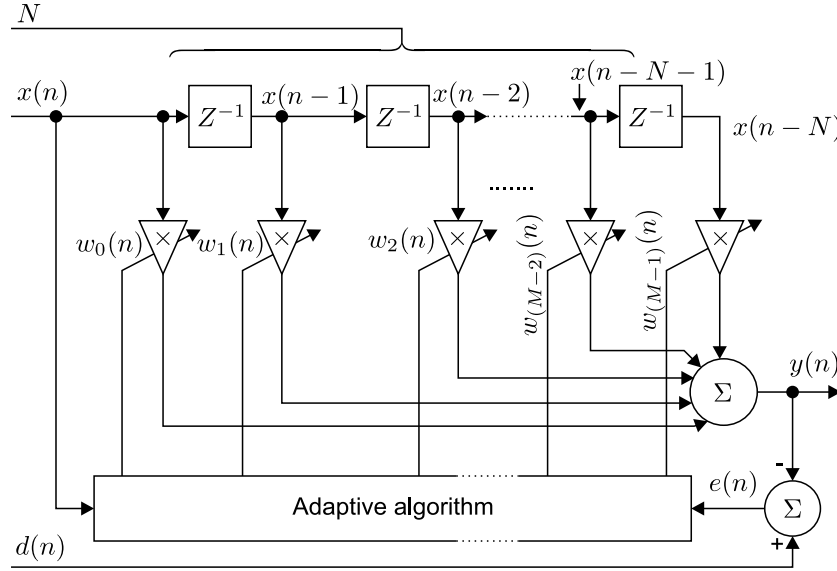
#### 3.1 LMS a RLS algoritmus

Pro potlačení šumu v řečových signálech se mezi nejpožívanější metody řadí adaptivní systémy založené na algoritmech LMS (Least Mean Squares) a RLS (Recursive Least Square).

Algoritmus LMS představuje zástupce stochastických gradientních algoritmů, založených na teorii Wienerovy filtrace, stochastickém průměrování a metodě nejmenších čtverců. Algoritmus



RLS patří do skupiny rekursivních algoritmů, založených na teorii Kalmanovy filtrace, časovém průměrování a metodě nejmenších čtverců. Struktura filtru je u obou algoritmů stejná, jedná se o lineární FIR filtr s transverzální strukturou, kde  $N$  značí jeho řád [17, 21, 22].



Obrázek 9: Blokové schéma adaptivního filtru [28]

$\mathbf{x}(n)$  představuje vektor vstupního signálu,  $\mathbf{w}(n)$  váhy filtru,  $z^{-1}$  zpoždění,  $y(n)$  výstupní signál filtru,  $d(n)$  požadovaný signál a  $e(n)$  odchylku (rozdíl) požadované a výstupní hodnoty.

U LMS platí, že výstup filtru je dán sumou součinu vektoru vstupních vzorků  $\mathbf{x}(n)$  a vah transverzálního filtru  $\mathbf{w}(n)$  [17]:

$$y(n) = \sum_{i=0}^{N-1} \mathbf{w}(n) \mathbf{x}(n-i) \quad (3.1)$$

Chyba  $e(n)$  se vypočte jako rozdíl mezi výstupní hodnotou  $y(n)$  a požadovanou hodnotou  $d(n)$ . Následně jsou váhy filtru přepočteny podle rovnice [17]:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + 2\mu e(n) \mathbf{x}(n), \quad (3.2)$$

kde  $\mu$  označuje velikost kroku, a jedná se o malou kladnou hodnotu.

RLS počítá výstup jako součin vah filtru  $\mathbf{w}(n)$  z předchozí iterace a současného vstupního vektoru  $\mathbf{x}(n)$  [17]:

$$\bar{y}_{n-1}(n) = \bar{\mathbf{w}}^T(n-1) \mathbf{x}(n) \quad (3.3)$$

Váhy filtru jsou aktualizovány podle rovnice [17]:

$$\mathbf{w}(n) = \bar{\mathbf{w}}^T(n-1) + \mathbf{k}(n)\bar{e}_{n-1}(n), \quad (3.4)$$

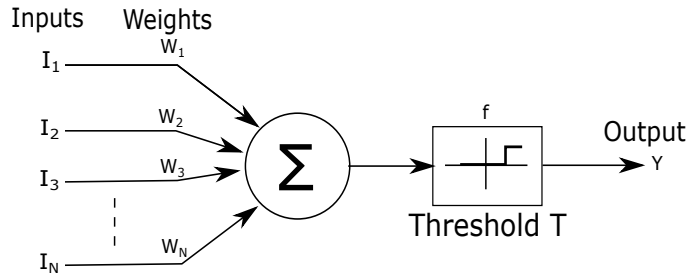
kde  $\mathbf{k}(n)$  představuje vektor zesílení.

Rozdíl mezi těmito algoritmy je ve vlastním statistickém pojetí. U RLS jsou počítány průměrné hodnoty veličin z časových vývoje, zatímco u LMS se jedná o průměr vzorků z několika realizací stejného náhodného procesu. RLS algoritmus má oproti LMS větší výpočetní náročnost, avšak rychlost konvergence je několikanásobně vyšší, což plyne z použitého časového průměrování [17, 21]. Podrobné odvození obou metod lze nalézt např. v [26].

### 3.2 Umělé neuronové sítě

Umělé neuronové sítě (ANN) jsou matematické modely, které svým chováním napodobují mozkové neurony. Základním stavebním prvkem je, stejně jako u lidské neuronové sítě, neuron. Tento neuron je schopen přijímat vnější signály, transformovat je a vyslat výstupní signál. Počet vstupních signálů do neuronu není omezen, přičemž jeden z nich určuje tzv. prahovou hodnotu  $\theta$ . Neurony jsou vzájemně propojeny synapsemi, pomocí kterých probíhá přenos signálu. Každá synapse disponuje vlastní vahou  $w$ , která pozměňuje procházející signál.

Základní model neuronu definoval McCulloch s Pittsem. Ten je tvořen pevným počtem vstupů, konstantní prahovou hodnotou, a jedním binárním výstupem. Síť, která obsahuje pouze jeden neuron typu McCulloch je označována jako perceptron [25].



Obrázek 10: Základní model neuronu [25]

Vstupní signál do neuronu je dán sumou jednotlivých signálů  $I_i$ , které jsou vynásobeny vahami synapsí  $w_i$ . Pokud potenciál (vážený součet vstupních signálů) neuronu překoná prahovou hodnotu, neuron se považuje za aktivní (dojde k excitaci na hodnotu 1), v opačném případě je neuron pasivní (excitace na hodnotu 0). Výstupní signál (excitovaná hodnota) vzniká transformací vstupního signálu přes aktivační (někdy též označovaná jako přenosová) funkci  $f$  [24].

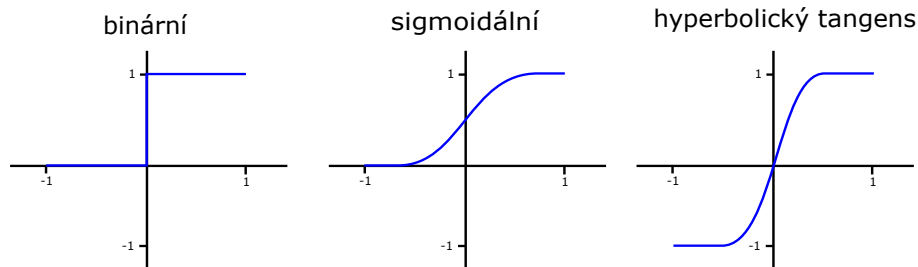
$$y = f\left(\sum_{i=1}^N I_i W_i + \theta\right) \quad (3.5)$$

Každý neuron má vlastní aktivační funkci  $f$ . Aktivační (přenosová) funkce je funkce, která transformuje vstupní signál neuronu na výstupní v intervalech 0 až 1 a -1 až +1. Tato funkce může být skoková nebo spojitá a musí být monotónní. Příkladem aktivační funkce může být funkce binární, sigmoidální nebo hyperbolický tangens [23].

$$f_{binarni}(x) = \begin{cases} 0 & x < \theta \\ 1 & x > \theta \end{cases}, \quad (3.6)$$

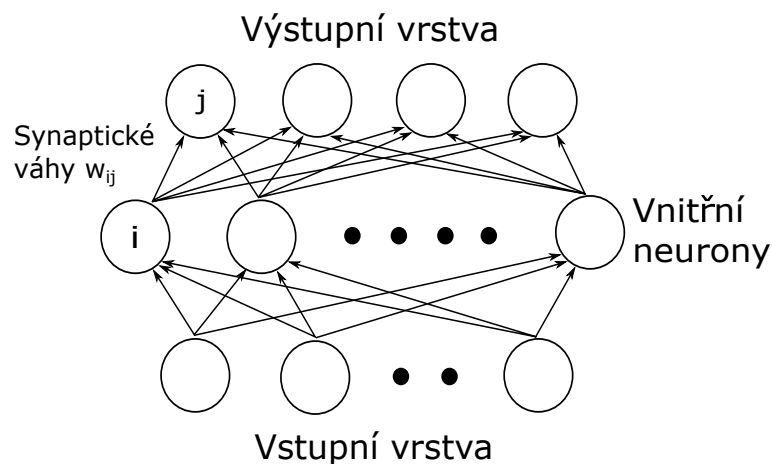
$$f_{sigmoid}(x) = \frac{1}{1 + e^{-x}} \quad (3.7)$$

$$f_{hyperTan}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (3.8)$$



Obrázek 11: Typy aktivačních funkcí

V neuronových sítích jsou neurony uspořádány do vrstev. Podle počtu vrstev v síti pak rozlišujeme síť jednovrstvé (Hopfieldova síť, Kohonenova síť) a vícevrstvé (Klasická vícevrstvá síť s algoritmem Backpropagation). Vrstvy jsou rozděleny na vstupní, výstupní a skryté, přičemž každý neuron v dané vrstvě je synapsemi spojen se všemi neurony v následující vrstvě. Pro vícevrstvé síť platí, že počet vrstev je alespoň tři, maximální počet však omezen není [23].



Obrázek 12: Schéma vícevrstvé NN [24]

Informace je v neuronové síti nejprve zpracovávána v tzv. dopředném chodu (ang. feedforward). Tuto fázi nazýváme jako aktivační (vybavovací) proces. Algoritmus dopředného šíření lze popsat takto [24]:

1. Excitace neuronů vstupní vrstvy na úrovni 0 až 1.
2. Excitace jsou synapsemi přivedeny do následující vrstvy a upraveny pomocí vah synapsí.
3. Provedení sumace příchozích upravených signálů v následující vrstvě a excitace na úrovni danou aktivační funkcí neuronu.
4. Postupné šíření skrz všechny vrstvy až k vrstvě výstupní, kde získáme excitační stavy všech jejích neuronů.

Stěžejní vlastností neuronových sítí je schopnost učení, která má za cíl upravit hodnoty vah synapsí  $w_i$  tak, aby bylo dosaženo požadovaných výstupů. Tento proces lze rozdělit na dva typy: s učitelem (sít s algoritmem BackPropagation) a bez učitele (Hopfieldova síť). Učení s učitelem znamená, že se síť snaží přizpůsobit svou odezvu na vstupní informace tak, aby se její momentální výstup co nejvíce podobal požadovanému originálu. Učení bez učitele je proces, ve kterém síť vychází z informací, které jsou obsaženy ve vstupních vektorech [23, 24].

Pro správné naučení je potřeba stanovit tzv. trénovací množinu a vhodnou metodu, která je schopna nastavit váhy synapsí na základě prvků v trénovací množině. Trénovací množinu můžeme definovat jako uspořádané dvojice prvků vektorů vstupních hodnot a požadovaných výstupních hodnot [23, 24].

Nejrozšířenější metodou adaptace sítě je zmíněný algoritmus zpětného chodu (ang. BackPropagation). Tuto fázi nazýváme jako adaptační (učicí) proces. Tento algoritmus lze popsat následovně [24]:

1. Určíme excitaci výstupní vrstvy neuronů na základě dopředného šíření prvku  $I_i$  trénovací množiny.
2. Porovnáme skutečnou odezvu NN s požadovanou odezvou prvku  $O_i$ .
3. Rozdíl mezi skutečnou a požadovanou odezvou definuje chybu neuronové sítě. Tuto chybu pak v určitém poměru - learning rate - "vracíme zpět" do neuronové sítě formou úpravy synaptických vah mezi jednotlivými vrstvami směrem od výstupní vrstvy k vrstvě vstupní.
4. Po průchodu celé trénovací množiny sítí (jedna epocha) je určena globální chyba. Pokud je její hodnota vyšší než požadovaná, proces se opakuje.

Využití neuronových sítí je velmi široké, lze je využít např. pro klasifikaci, filtraci (lineární i nelineární filtry), predikci chování, optimalizaci nebo identifikaci signálů [23].

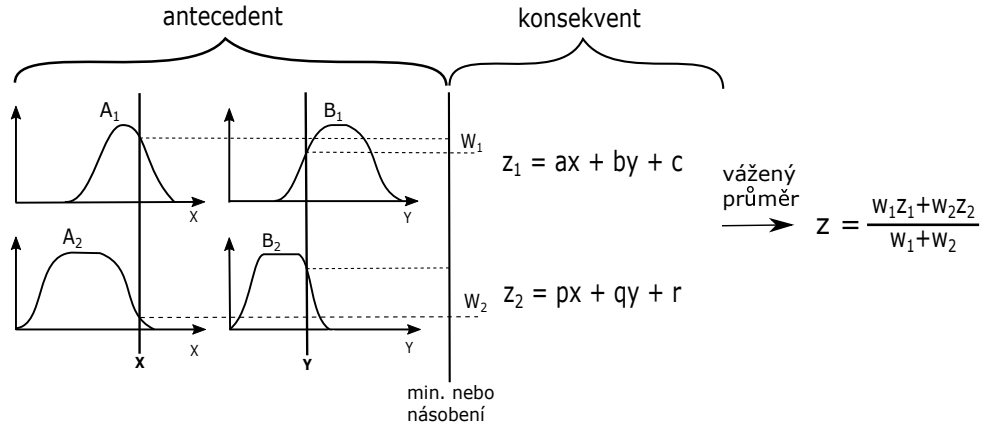
## 4 ANFIS

Systém ANFIS poprvé představil ve své studii [1] autor J.-S. R. Jang v roce 1993. Autor ve svém článku představuje nový systém pro využití v oblasti modelování nelineárních funkcí v měřicích a řídicích systémech. Představovaný systém je adaptivní dopřednou neuronovou sítí, která je funkčně ekvivaletní fuzzy inferenčním systému (FIS) typu Takagi-Sugeno s dvěma vstupy a jedním výstupem. ANFIS je schopen zkombinovat principy fuzzy logiky, která zahrnuje lidské znalosti v podobě tzv. podmíněných pravidel, a neuronových sítí, které systému dávají schopnost adaptace. Systém pro nalezení optimálních podmíněných pravidel využívá hybridní metodu učení, která spočívá v kombinaci algoritmu zpětného šíření (BackPropagation) pro vstupní parametry a metody nejmenších čtverců pro parametry výstupní. Výhoda ANFIS spočívá v rychlosti. Ve srovnání s klasickými neuronovými sítěmi je zde pouze malý počet parametrů, které podléhají učení. Rychlost je také dosažena využitím hybridního učení, kde je velká část parametrů učená algoritmem LMS, díky čemuž klesá výpočetní náročnost způsobená algoritmem zpětného šíření [1, 4].

Správně funkční systém vyžaduje následující požadavky [26]:

- jedná se o Takagi-Sugeno systém prvního nebo nultého řádu,
- systém má jeden výstup,
- všechny výstupní funkce příslušnosti patří ke stejnému typu, tzn., jsou lineární nebo konstantní,
- nejsou sdílena pravidla,
- různá pravidla nemohou sdílet různé výstupní funkce příslušnosti,
- počet funkcí příslušností výstupu je stejný jako počet pravidel.

ANFIS je založen na konstrukci IF-THEN pravidel typu Takagi-Sugeno, který je typický tím, že model konsekventu je realizován ve formě lineární kombinace vstupních proměnných (viz. obr. 13). Výstupem modelu využívající Takagi-Sugeno pravidla je přímo ostrá hodnota, není tedy nutné dělat defuzzifikaci. Využití tohoto modelu je výhodné právě u nelineárních soustav, u kterých je známo, že je lze částečně linearizovat, ačkoliv oblasti linearizace nemůžeme přesně stanovit [26].



Obrázek 13: Mechanismus rozhodování dle modelu Takagi-Sugeno. Upraveno z [1]

Pravidla typu Takagi-Sugeno jsou zapsána ve formě [1, 2, 4]:

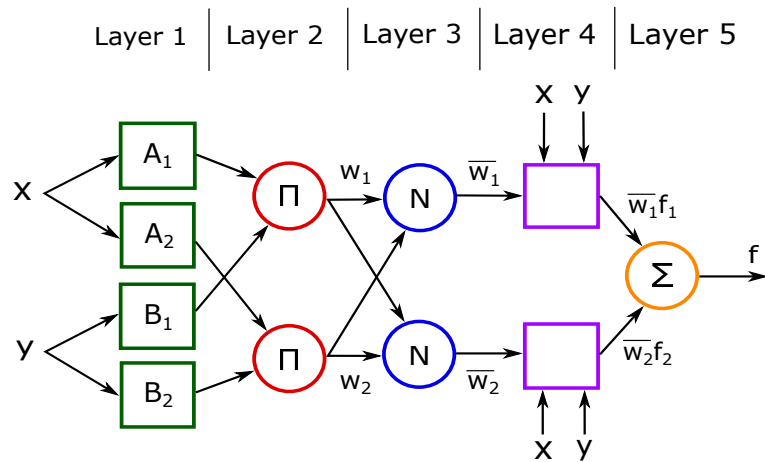
$$\text{IF } x \text{ is } A_1 \text{ and } y \text{ is } B_1, \text{ THEN } f_1 = p_1x + q_1y + r_1, \quad (4.1)$$

$$\text{IF } x \text{ is } A_2 \text{ and } y \text{ is } B_2, \text{ THEN } f_2 = p_2x + q_2y + r_2, \quad (4.2)$$

kde  $x$  a  $y$  představují vstupní hodnoty,  $A, B$  jsou hodnotami v intervalu  $\langle 0,1 \rangle$ ,  $f_1$  a  $f_2$  jsou výstupy filtru a  $p, q, r$  jsou parametry výstupní lineární funkce [1, 2, 4].

#### 4.1 Architektura

Původní architektura ANFIS je zobrazena na obr. 14. Jang ji ve své studii popisuje jako síť, která je složena z pěti vrstev s dvěma vstupy a jedním výstupem. Vrstva označená čtvercem má adaptivní parametry a vrstvy označené kruhem pevné parametry. Rozdílem oproti klasické neuronové síti je zde to, že synapse nedisponují žádnými vahami, znázorňují tak pouze směr toku dat skrz síť [1].



Obrázek 14: Struktura ANFIS. Upraveno z [20]

Matematický model vrstev je popsán pomocí následujících rovnic [1]:

První vrstva je použita pro fuzzifikaci vstupních hodnot. Přenosové funkce neuronů zde představují funkce příslušnosti. Při učení jsou v této vrstvě nastavovány parametry funkcí příslušnosti, které tvoří antecedent pravidel.

$$O_{1,i} = \mu A_i(x), \text{ pro } i = 1, 2, \quad (4.3)$$

$$O_{1,i} = \mu B_{i-2}(y), \text{ pro } i = 3, 4, \quad (4.4)$$

kde  $x$  a  $y$  jsou vstupy do uzlu,  $A_i$  a  $B_{i-2}$  je fuzzy set a  $\mu$  představuje funkci příslušnosti. Funkce příslušnosti může být funkce Gaussovská, zvonová, triangulární nebo trapezoidální. Předpis pro zvonovou funkci:

$$\mu A_i(x) = \frac{1}{1 + \left[ \left( \frac{x-c_i}{a_i} \right)^2 \right] b_i}, \quad (4.5)$$

kde  $a_i$ ,  $b_i$  a  $c_i$  jsou parametry udávající tvar funkce.

Druhá vrstva je označena  $\Pi$  a výstupem každého uzlu je součin všech jeho vstupních signálů. Vstupem do uzlů jsou signály z předchozí vrstvy a každý uzel zde odpovídá jednomu fuzzy pravidlu. Výstupem je tedy vypočtená síla pravidla.

$$w_i = \mu A_i(x) \mu B_i(y), \quad i = 1, 2 \quad (4.6)$$

Třetí vrstva je označena písmenem  $N$ . Výstupy uzlů vyjadřují poměr váhy jednotlivých pravidel k součtu vah všech pravidel  $w$ . Počet uzlů je totožný s předcházející vrstvou.

$$\overline{w}_i = \frac{w_i}{w_1 + w_2}, \quad i = 1, 2 \quad (4.7)$$

Ve čtvrté vrstvě je každý uzel adaptivní s přenosovou funkcí odpovídající požadovanému tvaru konsekventu (konstantní nebo lineární).

$$O_{4,i} = \overline{w}_i f_i = \overline{w}_i (p_i x + q_i y + r_i), \quad (4.8)$$

kde  $\overline{w}_i$  normalizovaná síla pravidla z předchozí vrstvy a  $p$ ,  $q$ ,  $r$  jsou parametry konsekventu.

Poslední, výstupní, vrstvou je vrstva pátá, označená  $\Sigma$ . V této vrstvě dochází k výpočtu celkového výstupu všech příchozích signálů ze čtvrté vrstvy.

$$O_{5,i} = \text{overall output} = \sum w_i f_i = \frac{\sum w_i f_i}{\sum w_i} \quad (4.9)$$

## 4.2 Učení

Použití tradiční metody sestupného gradientu (GD) sebou přináší nenulovou pravděpodobnost, že dojde k uváznutí v lokálním minimu účelové funkce. V naprosté většině aplikací využívá ANFIS hybridní učící algoritmus. Tento algoritmus představuje kombinaci algoritmu zpětného šíření s metodou sestupného gradientu (BP/GD) a metody nejmenších čtverců (LMS). Ve fázi učení dochází k úpravě parametrů antecedentu a konsekventu IF-THEN pravidel. Při dopředném chodu informace sítě se uplatňuje LMS, a to tak že pozměňuje parametry konsekventu (koeficienty lineární funkce), zatímco parametry antecedentu zůstávají pevné. Naopak při zpětném chodu je aplikován algoritmus BP/GD, který upravuje parametry antecedentu (funkce příslušnosti vstupních proměnných), zatímco parametry konsekventu jsou fixní [1, 26].

### 4.2.1 Dopředný chod

Při dopředném chodu síť pracuje s metodou nejmenších čtverců pro určení parametrů lineární funkce konsekventu. Výstupní lineární funkce lze obecně zapsat do tvaru:

$$Z_i = p_i X + q_i Y + R_i \quad (4.10)$$

Na vstup ANFIS jsou přivedeny tréninkové vzory a postupně vypočteny výstupy jednotlivých vrstev a identifikovány parametry konsekventu odhadem pomocí LMS. Předpokládáme-li  $m$  počet tréninkových vzorů, pak lze sestavit  $m$  lineárních rovnic zahrnující parametry  $p$ ,  $q$  a  $r$ . Následující popis lze nalézt např. v [1] nebo [26].

$$\left\{ \begin{array}{l} Z(1) = \overline{w}_1(1)f_1(1) + \dots + \overline{w}_n(1)f_n(1) \\ Z(2) = \overline{w}_2(2)f_2(2) + \dots + \overline{w}_n(2)f_n(2) \\ Z(m) = \overline{w}_m(m)f_m(m) + \dots + \overline{w}_n(m)f_n(m) \end{array} \right\} \quad (4.11)$$

kde  $m$  je počet tréninkových vzorů,  $n$  počet uzlů v pravidlové vrstvě a  $X$ ,  $Y$  a  $Z$  (viz. rovnice 4.10) jsou vstupní/výstupní vzory.

Tyto rovnice lze zapsat do tvaru:

$$Z = A\mathbf{k}, \quad (4.12)$$

kde  $Z$  je matice  $m \times 1$  výstupního vektoru ve tvaru:

$$\left\{ \begin{array}{l} Z(1) \\ Z(2) \\ Z(m) \end{array} \right\} \quad (4.13)$$



$A$  je matice  $m \times n$ :

$$\begin{pmatrix} \bar{w}(1) & \bar{w}(1)X(1) & \bar{w}(1)Y(1) & \dots & \bar{w}_n(1) & \bar{w}_n(1)X(1) & \bar{w}_n(1)Y(1) \\ \bar{w}(2) & \bar{w}(2)X(2) & \bar{w}(2)Y(2) & \dots & \bar{w}_n(2) & \bar{w}_n(2)X(2) & \bar{w}_n(2)Y(2) \\ \bar{w}(m) & \bar{w}(m)X(m) & \bar{w}(m)Y(m) & \dots & \bar{w}_n(m) & \bar{w}_n(m)X(m) & \bar{w}_n(m)Y(m) \end{pmatrix} \quad (4.14)$$

a  $\mathbf{k}$  je vektor o velikosti  $n \times 1$  neznámých parametrů  $p$ ,  $q$  a  $r$ :

$$\mathbf{k} = [p_1 \quad q_1 \quad r_1 \quad \dots \quad p_n \quad q_n \quad r_n]^T \quad (4.15)$$

V případě, že počet vstupně/výstupních vzorů  $m$  převyšuje nad počtem neznámých konsekventů  $n$ , dochází k tzv. přeučení problému, který nemusí mít žádné řešení. V takovém případě pak hledáme odhad  $\mathbf{k}$  metodou LMS, který označujeme jako  $\mathbf{k}^*$ . Tento odhad minimalizuje chybu  $\|A\mathbf{k} - Z\|^2$  užitím pseudo-inverzní metody.

$$\mathbf{k}^* = (A^T A)^{-1} A^T Z, \quad (4.16)$$

kde  $T$  značí transponovanou matici a výpočet  $(A^T A)^{-1} A^T$  psoudo-inverze  $A$  v případě, že  $(A^T A)$  je nesingulární.

Výpočtem neznámých parametrů konsekventu lze stanovit celkový výstup ANFIS  $Z_{aktualni}$  a chybový vektor  $\mathbf{e}$ .

$$\mathbf{e} = Z - Z_{aktualni} \quad (4.17)$$

#### 4.2.2 Zpětný chod

Ve zpětném chodu se uplatňuje algoritmus BP/GD. Stanovený chybový signál  $e$  z předchozího dopředného chodu je distribuován sítí zpět za pomoci tzv. řetězového pravidla (derivace složené funkce). Tímto způsobem dochází k úpravě parametrů v antecendentu a tvaru funkce příslušnosti (viz. obr. 15). Jako příklad zde bude uveden přepočtení parametru  $a$  pro zvonovou funkci příslušnosti. Popis algoritmu byl převzat z [26].

$$\Delta a = -\eta \frac{\partial E}{\partial a} = -\eta \frac{\partial E}{\partial e} \times \frac{\partial e}{\partial Z_{aktualni}} \times \frac{\partial Z_{aktualni}}{\partial (\bar{w}_i f_i)} \times \frac{\partial \bar{w}_i f_i}{\partial \bar{w}_i} \times \frac{\partial \bar{w}_i}{\partial w_{A_i}} \times \frac{\partial w_{A_i}}{\partial a}, \quad (4.18)$$

kde  $\eta$  je učicí koeficient a  $E$  je aktuální chyba na výstupním neuronu.

$$E = \frac{1}{2} e^2 = \frac{1}{2} (Z - Z_{aktualni})^2 \quad (4.19)$$

Platí také, že:

$$\Delta a = -\eta(Z - Z_{aktualni})(-1)f_1 \times \frac{\bar{w}_i(1 - w_i)}{w_i} \times \frac{w_i}{w_{Ai}} \times \frac{\partial w_{Ai}}{w_a} \quad (4.20)$$

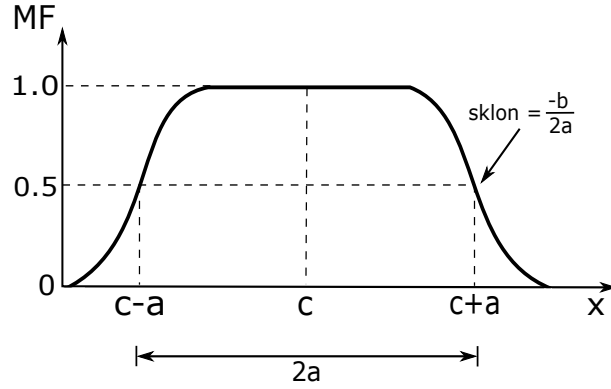
další možný tvar:

$$\Delta a = -\eta(Z - Z_{aktualni})f_1\bar{w}_i(1 - \bar{w}_i) \times \frac{1}{w_{Ai}} \times \frac{\partial w_{Ai}}{w_a} \quad (4.21)$$

pro zvonovou funkci bude tedy platit:

$$\frac{\partial w_{Ai}}{\partial a} = \frac{1}{\left[1 + \left(\frac{X-a}{c}\right)^{2b}\right]^2} \times \frac{1}{e^{2b}} \times 2b \times (X - a)^{2b-1} \times (-1) = w_{Ai}^2 \times \frac{2b}{c} \times \left(\frac{X - a}{c}\right)^{2b-1} \quad (4.22)$$

Koeficienty  $b$  a  $c$  získáme obdobně.



Obrázek 15: Koeficienty definující tvar zvonové funkce. Upraveno z [1]

## 5 Využití ANFIS ve zpracování řeči

ANFIS lze využít v mnoha oblastech automatického řízení, rozpoznávání vzorů, předvídání chování nebo modelování systémů. V případě, kdy počet IF-THEN pravidel není omezen, Takagi-Sugeno systém nultého řádu dokáže aproximovat jakoukoliv nelineární funkci. Můžeme tedy ANFIS považovat za univerzální estimátor [30].

V následujících kapitolách budou uvedeny studie, zabývající se využitím ANFIS v oblasti zpracování a zkvalitnění signálů.

### 5.1 Potlačení šumu v nahrávce řeči

Ve studii z roku 2015 od autorů R. Martínka a kol. [2] je ANFIS využit pro potlačení šumu v řeči. V místnosti byly umístěny dva zdroje zvuku. Prvním byl reproduktor produkující pouze šum, dosahující intenzity 80 – 100 dB. Druhým byl zdroj lidské řeči o intenzitě 30 – 40 dB. Systém má dva vstupy v podobě referenčního mikrofону, snímající pouze šum, a primárního mikrofону snímající řeč, ale také přičtený šum z místnosti. Signál z primárního mikrofону odpovídá součtu užitečného signálu a deformovaného šumu, který vznikl nelineární transformací původního šumu.

Trénování neuronové sítě proběhlo podle hybridního algoritmu kombinující RLSE (Recursive Least Squares Estimation) a BP/GD (BackPropagation Gradient Descent). V experimentu byla vypočtena hodnota SSNR (Segmental Signal to Noise Ratio) a DTW (Dynamic Time Warping). Hodnota SSNR byla vypočtena před a po průchodu signálu skrz ANFIS. DTW kritérium hodnotí vzdálenosti dvou vektorů užitečného signálu [2].

Autoři článku v rámci experimentů otestovali několik variant systému ANFIS. Tyto varianty se lišily v počtu uzlů sítě nebo v počtu lineárních či nelineárních parametrů (modely A, B, C a D). Z dosažených výsledků je patrné, že ANFIS dosahuje lepších výsledků než porovnávané adaptivní algoritmy LMS a RLS. Ze všech typů zkoumaných ANFIS autoři vyhodnotili model B jako nejvíce vhodný z hlediska matematické náročnosti a dosažených výsledků. Z hlediska kvality výstupního signálu si model A vedl nejhůře, modely B, C a D dosáhly přibližně stejných výsledků [2].

Obdobnou studii zabývající se odšuměním řečové nahrávky provedli v roce 2010 M. H. Miry a kol. [6], Thevaril a Kwan [8] v roce 2005 a J. Kumar a kol. [3] v roce 2015. Cílem studií bylo potlačení šumu v nahrávce řeči a dosažení co nejvyšší hodnoty SNR. Všichni autoři dospěli ke stejným závěrům jako u [2]. Proběhlo také porovnání s algoritmem LMS, který dosáhl nižší hodnoty SNR i pomalejší rychlosti filtrace než systém ANFIS [3, 6].

Studie provedená autorem M. H. Miry byla také zaměřena na porovnání efektivity různých tvarů funkce příslušnosti. Pro měření byly vygenerovány tři druhy Gaussova šumu. Jejich hodnoty SNR (signal to noise ratio) byly -4 dB, 0 dB a 4 dB. Každý ze tří druhů šumu byl testován na různých druzích MF a pro každý bylo vypočteno SNR.

Provedené experimenty ukázaly, že rozdíly mezi jednotlivými typy funkcí nejsou velké. Nejlepší hodnoty SNR dosáhl trojúhelníkový tvar funkce příslušnosti s hodnotou  $\text{SNR} = 10,3 \text{ dB}$  [6].

Studie autora J. Kumar oproti ostatním testovala různé typy rušení (stacionární i nestacionární). Pro testování byly vytvořeny čtyři typy signálu, obsahující užitečný signál plus jednotlivé rušení. Specifické nastavení ANFIS autoři uvádějí pouze u měření se stacionárním typem šumu, který představoval zvuk běžícího motoru. Pro tento signál byl nastaven počet uzlů na 20, počet lineárních parametrů 82, počet nelineárních parametrů 48, počet testovacích párů dat 860 a počet fuzzy pravidel 16. Autor článku poukazuje na úspěšné odrušení všech typů šumu v krátkém čase [3].

## 5.2 Porovnání s metodami HHT a WT

Autoři L. Yang, H. Yang a Shen provedli v roce 2010 studie [5, 7] zaměřené na porovnání efektivity ANFIS s metodami HHT (Hilbert-Huang Transform) a WT (Wavelet Transform). Pro obě studie byly nasimulovány dva informační signály. Pro tyto signály byl vytvořen náhodný šumový signál s normálním rozložením. Následně byl tento signál modifikován nelinearitou a přičten k informačnímu signálu. Oba měřené signály byly zpracovány metodami HHT a WT a následně pomocí ANFIS.

U obou testů autoři došli k závěru, že ANFIS odfiltroval šum lépe v oblastech vyšších frekvencí, zatímco v oblastech nižších frekvencí si vedla nejlépe metoda vlnkové transformace. V oblastech vyšších frekvencí překrývá šum výrazně frekvenční pásmo signálu, čímž u daných metod dochází ke ztrátám informace [5, 7]. Optimální metodu filtrace hodnotí autoři článku kombinací ANFIS a WT, která vede k lepším výsledkům [7].

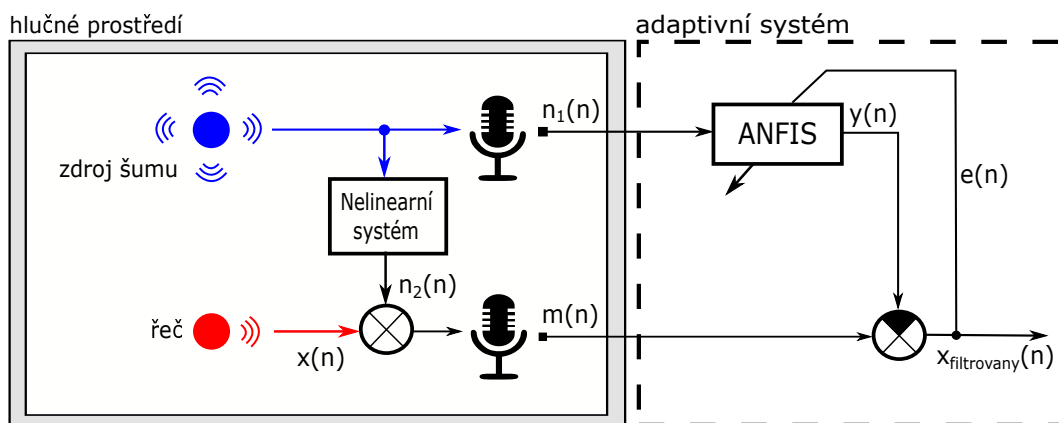
## 6 Implementace systému ANFIS

Praktická část této diplomové práce je věnována vytvoření vlastního adaptivního systému a provedení experimentů v oblasti potlačení šumu v řečových nahrávkách. Tato kapitola nabídne popis návrhu a implementace řešení v programovacím jazyce C#. Při implementaci byla využita již existující knihovna funkcí ANFIS.

### 6.1 Architektura a návrh systému

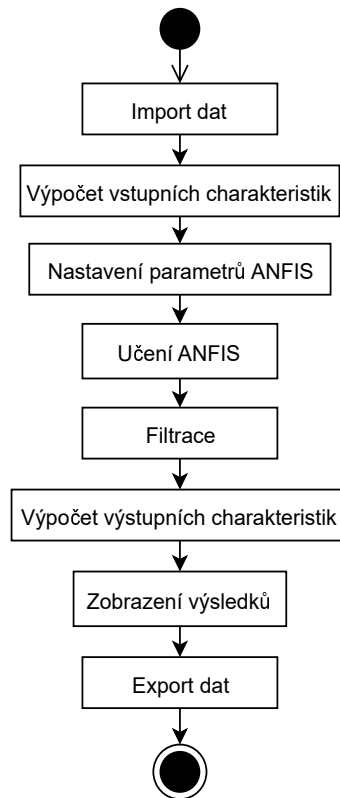
Na obr. 6.1 je znázorněn navržený adaptivní systém pro zpracování zašuměného řečového signálu. Adaptivní systém disponuje dvěma vstupy v podobě řeči, zaznamenané primárním mikrofonom, a šumu, zaznamenaný referenčním mikrofonom. Vstup označený  $n_1(n)$  je šum měřený přímo u zdroje. Druhý vstup  $m(n)$  je řeč sečtená se signálem  $n_2(n)$ , který představuje šum  $n_1(n)$  transformovaný neznámým nelineárním systémem prostředí.

Přímým vstupem do ANFIS jsou signály  $n_1(n)$  a  $n_1(n - 1)$ . Po každé iteraci je výstupní signál  $y(n)$  odečten od řečového signálu  $m(n)$ , čímž získáme chybový signál  $e(n)$ , který slouží pro zpětnou úpravu parametrů struktury algoritmem učení. Po ukončení všech iterací je výsledný signál  $x_{filtrovaný}(n)$  získán odečtením  $m(n)$  a  $y(n)$ .



Obrázek 16: Schéma adaptivního systému

Výše uvedený systém byl vytvořen jako desktopová aplikace s grafickým uživatelským rozhraním (viz. kapitola 6.1.1). V jádru aplikace je využit framework ANFIS, kolem kterého byl vytvořen adaptivní systém, umožňující zpracovat nahrané signály. Proces zpracování signálů je rozdělen do několika fází, které jsou zobrazeny na následujícím diagramu.



Obrázek 17: Diagram aktivit

Adaptivní systém pro svou správnou funkci vyžaduje načtení alespoň dvou signálů (zašuměný signál a šum), zatímco třetí (čistý) signál povinný není. Data mohou být do aplikace nahrána ve formě CSV nebo WAV souboru. Při načtení jsou zkontrolovány délky signálů. Pokud dojde k neshodě, nové data nebudou nahrána. Následně je provedena normalizace signálů podělením signálu maximální možnou hodnotou signálu (pro signál s hloubkou 16 bit je tato hodnota 32 768). V případě načtení čistého signálu jsou následně vypočteny vstupní hodnotící charakteristiky (SNR, SSNR, DTW a RMSE).

Uživatelským vstupem jsou parametry filtrace, kterými jsou počet trénovacích vzorků, počet pravidel, počet epoch a typ funkce příslušnosti. Pro správnou funkci systému je nutné, aby počet trénovacích prvků byl větší než počet pravidel.

Samotná filtrace neprobíhá v reálném čase, ale ve dvou fázích. V první dochází pouze k učení ANFIS. Nejprve jsou stanoveny počáteční IF-THEN pravidla na základě shlukování vstupně/výstupních dat metodou nejbližšího souseda. Následně dochází k ladění parametrů pravidel pomocí algoritmu učení. Ve druhé fázi jsou již signály filtrovány skrze adaptivní strukturu s pevně nastavenými parametry. Z hlediska celkové doby zpracování dat je první fáze nejvíce výpočetně náročná.

Výsledek filtrace je uživateli nabídnut v podobě grafů a výsledných hodnot SNR, SSNR, DTW, RMSE a celkového času filtrace.

Kromě nahrání signálů a provedení filtrace, aplikace dále umožňuje následující funkcionality:

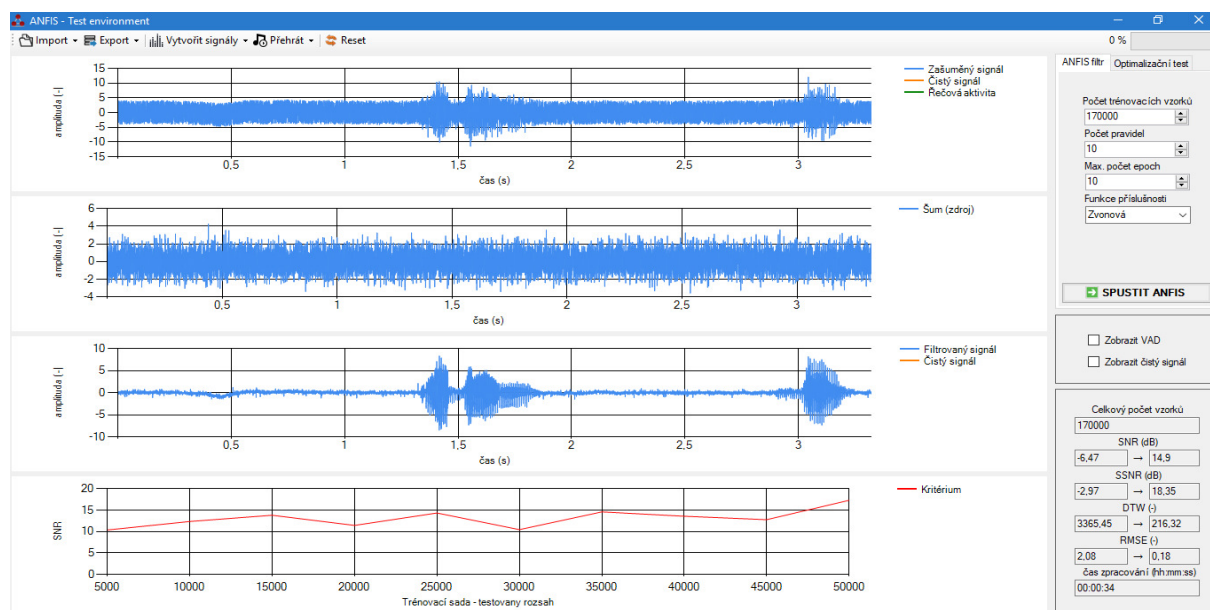
- Přehrání nahraných záznamů,
- oříznutí signálů,
- vytvoření umělého šumu (bílý šum),
- spuštění testu pro nalezení optimálních parametrů,
- export dat do CSV nebo WAV souboru.

Optimalizační test poskytuje znázornění závislosti dvou pozorovaných parametrů (např. závislost SSNR na počtu fuzzy pravidel). Uživatelským vstupem pro test jsou dva požadované parametry a nastavení počáteční a koncové hodnoty společně s krokem. Ostatní parametry jsou v testu fixní.

Architektura aplikace je rozdělena do dvou vrstev, prezentační a aplikační vrstvy.

### 6.1.1 Prezentační vrstva

Prezentační vrstva byla vytvořena pomocí knihovny tříd Windows Forms pro práci s grafickým uživatelským rozhraním (GUI). Toto prostředí umožňuje uživateli interakci s aplikací s výše uvedenými funkcionalitami.



Obrázek 18: Uživatelské prostředí aplikace

Hlavní součástí GUI jsou v centrální části umístěné grafy, zobrazující časové průběhy signálů. Horní graf poskytuje zobrazení zašuměné řeči společně s čistým signálem jako vizuální referenci. Druhý graf zobrazuje pouze šum, který společně se zašuměným signálem tvoří jeden ze vstupů do adaptivního systému. Třetí, výstupní, graf zobrazuje výsledek procesu filtrace. Poslední graf

slouží pro zobrazení výsledku optimalizačního testu, znázorňujícího závislost dvou pozorovaných parametrů.

Ovládání aplikace je soustředěno převážně do horního menu, obsahující prvky pro načtení signálů, export dat, vytvoření umělých signálů, přehrání záznamů, resetu aplikace a grafického zobrazení průběhu zpracování dat. Ovládání samotné filtrace je lokalizováno do pravé části, společně s nastavením parametrů filtrace a možností zobrazení čistého signálu a VAD.

### **6.1.2 Aplikační vrstva**

Aplikační vrstva se skládá ze dvou základních částí: ANFIS frameworku a implementovaného systému, který propojuje ANFIS s prezentační vrstvou. Jak již bylo zmíněno, aplikace byla vytvořena v programovacím jazyce C#, bylo tedy využito objektově orientované paradigma. Pro práci se zvukovými soubory (import a export) byla využita knihovna NAudio.

Veškeré procesy, s výjimkou optimalizačního testu, běží v hlavním vlákne aplikace. Optimalizační test běží jako paralelní proces. Pro tento proces je vytvořen počet vláken odpovídající počtu jednotlivých nastavení ANFIS (např. pro počet trénovacích vzorků od 10 do 100 s krokem 5 je počet vláken 20).



## 7 Experimenty

Tato kapitola bude věnována experimentům, které proběhly za účelem otestování vytvořené aplikace. Předmětem testování byla zejména kvalita filtrace a celkový čas, potřebný pro zpracování signálů adaptivním systémem. Kvalita filtrace byla posouzena s ohledem na nastavení parametrů ANFIS, jako je počet fuzzy IF-THEN pravidel, počet epoch a také tvar použité funkce příslušnosti. Počet trénovacích prvků byl nastaven vždy na celý signál.

Systém byl nejprve otestován na uměle vytvořených datech, která byla vygenerována samotnou aplikací, a poté na datech reálných, které představovaly zvukové nahrávky. Pro objektivní zhodnocení kvality filtrace byly vždy před a po průchodu dat adaptivním systémem vypočteny parametry SNR (odstup signálu od šumu), SSNR (segmentální odstup signálu od šumu), RMSE (relativní střední kvadratická odchylka) a DTW (dynamic time warping). Výpočet těchto charakteristik je podmíněn přítomností čistého, referenčního, signálu.

Pro lepší grafické zobrazení byly data z vytvořené aplikace nejprve exportována do prostředí MATLAB, a poté vykresleny do grafů.

### 7.1 Metody hodnocení kvality filtrace

Cílem adaptivního filtru je dosažení co nejmenší, ideálně nulové, amplitudy šumu ve zvukové stopě. Kvalitu filtrace lze posoudit jak subjektivními, tak objektivními metodami. Subjektivní metody jsou velmi závislé na posuzovateli a využijí se v případech, kde čistý, referenční, signál není k dispozici. Subjektivně tedy můžeme vlastním sluchem posoudit obsah šumu ve zvukové nahrávce. Na druhou stranu objektivní metody nejsou závislé na člověku a opírají se pouze o matematické rovnice. Předpokladem pro jejich stanovení je přítomnost referenčního signálu, od kterého se odvíjejí veškeré výpočty. V rámci této diplomové práce bude kvalita filtrace posuzována stanovením SNR, SSNR, RMSE a DTW.

#### 7.1.1 SNR a segmentální SNR

Parametr SNR, neboli odstup signálu od šumu, byl již zmíněn v kapitole 1.1.1. Tento parametr tedy vyjadřuje poměr výkonu čistého signálu ku výkonu šumu, který je k užitečnému signálu přičten (viz rovnice 1.4). SNR je nejprve vypočteno na vstupu systému (poměr zašuměného signálu a referenčního) a na výstupu systému (poměr filtrovaného signálu a referenčního). V aplikaci je parametr vypočten v jednotkách decibel.

$$\text{SNR}_{\text{dB}} = 20 \cdot \log\left(\frac{\text{signal}_{\text{RMS}}}{\text{sum}_{\text{RMS}}}\right) \quad (7.1)$$

Poměr výkonu je zde vyjádřen jako poměr jejich efektivních hodnot (Root Mean Square). RMS je pro sadu hodnot  $\{x_1, x_2, x_3, \dots, x_n\}$  vypočteno podle rovnice: CITACE

$$\text{RMS} = \sqrt{\frac{x_1^2 + x_2^2 + x_3^2 + \dots + x_n^2}{n}}, \quad (7.2)$$

kde  $x_i$  představuje jednotlivé prvky ve vektoru signálu a  $n$  vyjadřuje celkový počet prvků v signálu.

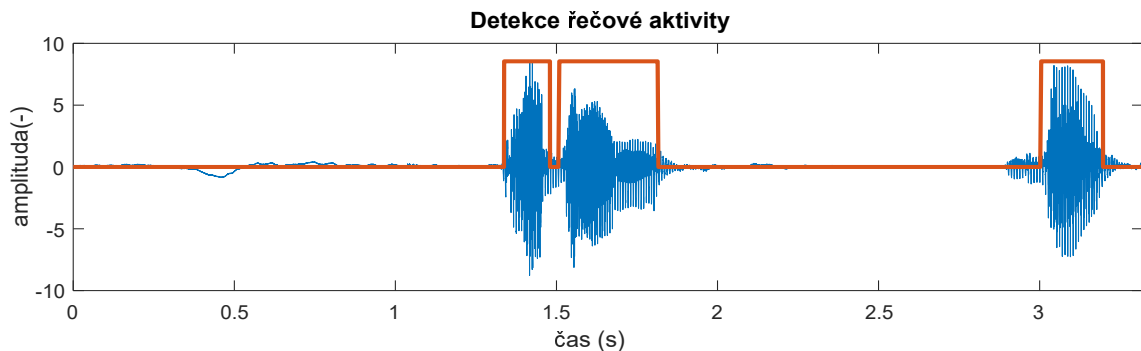
Odstup signálu od šumu vyjadřuje poměr signál/šum na celém rozsahu signálu. V oblasti zpracování řeči je však častěji využívaným parametrem SSNR, neboli segmentální odstup signálu od šumu. Díky tomuto jsme schopni stanovit poměr signál/šum pouze v oblastech našeho zájmu, tedy v oblastech řečové aktivity. Výpočet spočívá v rozdělení celého signálu na dílčí části (segmenty) a následné stanovení SNR v každém segmentu. [2, 29]

$$\text{SSNR} = \frac{1}{K} \sum_{i=0}^{L-1} \left( 10 \cdot \log \frac{\sum_{n=0}^{M-1} s_i^2(n)}{\sum_{n=0}^{M-1} n_i^2(n)} \cdot \text{VAD}_i \right), \quad (7.3)$$

kde  $K$  je počet segmentů obsahující řeč,  $L$  je celkový počet segmentů,  $M$  počet prvků v jednom segmentu a  $\text{VAD}_i$  představuje aktivitu řeči v  $i$ -tém segmentu a nabývá hodnot 0 (žádná aktivita) nebo 1 (řeč).

VAD (Voice Activity Detection) představuje klíčový parametr algoritmu a chyby při jeho určení resultují v chybu celého SSNR. VAD je nejčastěji určen na základě výpočtu energie signálu. Tento přístup však selhává v případech, kde je úroveň šumu velmi vysoká. [29]

V aplikaci je VAD vypočten pomocí čistého signálu. Jako první je určena obálka signálu za pomoci výpočtu absolutní hodnoty a následné aplikace dolnoprostopustního Butterworthova filtru. Z takto nalezené obálky signálu je vypočtena prahová hodnota, která byla experimentálně stanovena na 30% maximální amplitudy. Pokud vzorek signálu přesáhne tento práh, je detekován jako řeč. Na níže uvedeném obrázku je vidět jak byla aktivita řeči detekována aplikací, modrý signál představuje čistou řeč, oranžově je zobrazena řečová aktivita.



Obrázek 19: VAD v aplikaci

### 7.1.2 RMSE

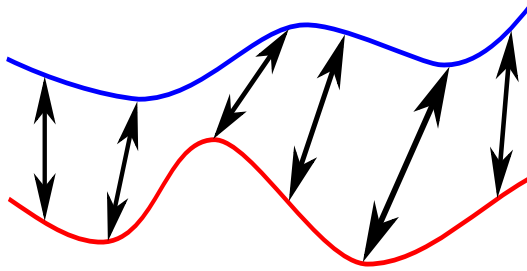
RMSE (Root Mean Square Error) vyjadřuje odchylku (rozptyl hodnot) od střední hodnoty, kterou v našem případě představuje požadovaný, referenční signál. RMSE je společně s jeho neodmocněnou formou MSE široce využíván ve statistice. MSE je definován jako suma kvadrátu rozdílu mezi referenčními a vstupními/výstupními vzorky signálu. Odmocněním MSE dostáváme předpis RMSE: [18]

$$\text{RMSE} = \frac{\sqrt{\sum_{i=0}^n (\text{signal}(i) - \text{ref}(i))^2}}{n} \quad (7.4)$$

Platí, že čím menší je hodnota RMSE resp. MSE, tím jsou si dva zkoumané signály podobnější.

### 7.1.3 DTW

Dynamic Time Warping je algoritmus, který měří podobnost dvou signálů, které se liší v rychlosti. DTW je často využíván v aplikacích rozpoznání řeči, právě z důvodu vyrovnání se s různou rychlostí mluvené řeči. DTW je schopno nalézt, na dvou časových řadách, body sobě odpovídající a určit jejich podobnost bez ohledu na jejich polohu způsobenou rozdílnou rychlostí obou signálů (viz obr. 20).



Obrázek 20: Dynamic Time Warping

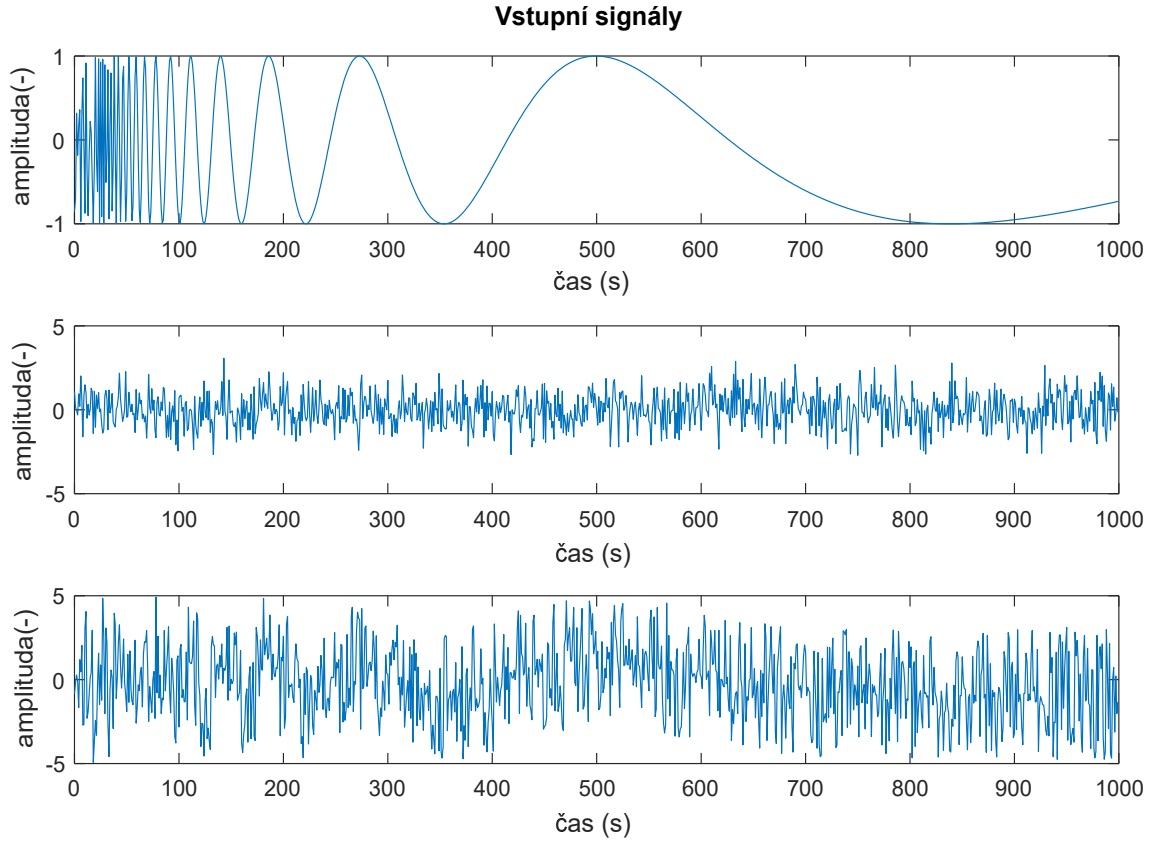
## 7.2 Testování na umělých datech

Prvotní otestování vytvořeného systému proběhlo pomocí uměle vytvořených dat. Tyto data představoval sinusový signál  $x(n)$  vytvořený podle rovnice 7.5 tak, aby demonstroval efektivitu ANFIS na různých frekvencích. Šum  $n_1(n)$  (zdrojový) představuje Gaussovský bílý šum, který byl následně zpožděn o jeden vzorek a transformován pomocí nelineární rovnice 7.6. Tento šum  $n_2(n)$  byl poté přičten k informačnímu signálu  $x(n)$ .

$$x(n) = \sin\left(\frac{40}{t + 0,1}\right) \quad (7.5)$$

$$n_2(n) = \frac{4 \cdot \sin(n_1(n)) \cdot n_1(n-1)}{1 + n_1(n-1)^2} \quad (7.6)$$

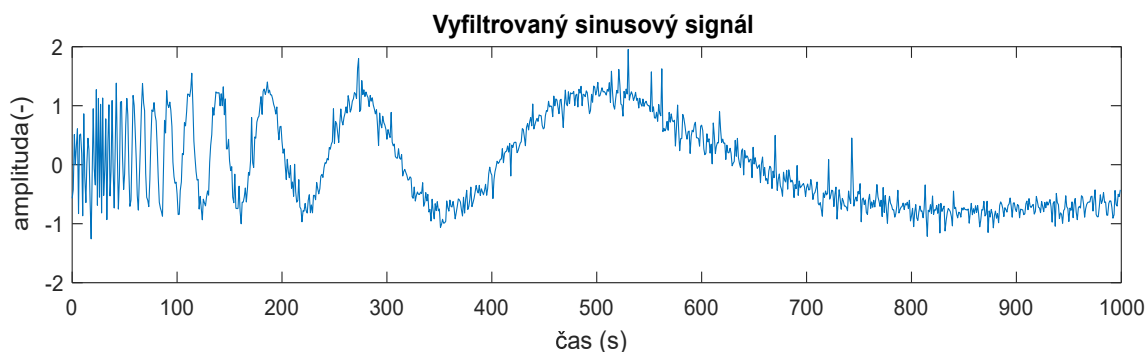
$$m(n) = x(n) + n_2(n) \quad (7.7)$$



Obrázek 21: Čistý sinusový signál (nahore), bílý šum (uprostřed) a zašuměný signál (dole)

Na obr. 21 jsou zobrazeny vytvořené signály. Na dolním grafu je vidět zašuměný signál  $m(n)$ , který byl vytvořen součtem signálů  $n_2(n)$  a  $x(n)$ . Úkolem ANFIS je nalezení nelineárního vztahu mezi  $n_1(n)$  a  $n_2(n)$ , který je požadovaným výstupem adaptivního systému. Pro správnou funkci ANFIS je potřeba stanovit vstupní a výstupní data pro trénování. ANFIS pracuje jako dopředná neuronová síť s učitelem, je tedy nutné jí poskytnout hledaný signál. Šum  $n_2(n)$  není přímo k dispozici, avšak můžeme předpokládat, že signál  $m(n)$  je jeho "kontaminovanou" verzí. Vstupními daty budou v našem případě zdrojový šum  $n_1(n)$  a zpožděný šum  $n_1(n - 1)$ .

Pro test byl nastaven počet testovacích vzorků na celý signál, tedy 1000, počet fuzzy pravidel na třicet, maximální počet epoch na padesát, a typ funkce příslušnosti na Gaussovskou. Vyfiltrovaný sinusový signál je uveden na obr. 22.



Obrázek 22: Vyfiltrovaný sinusový signál

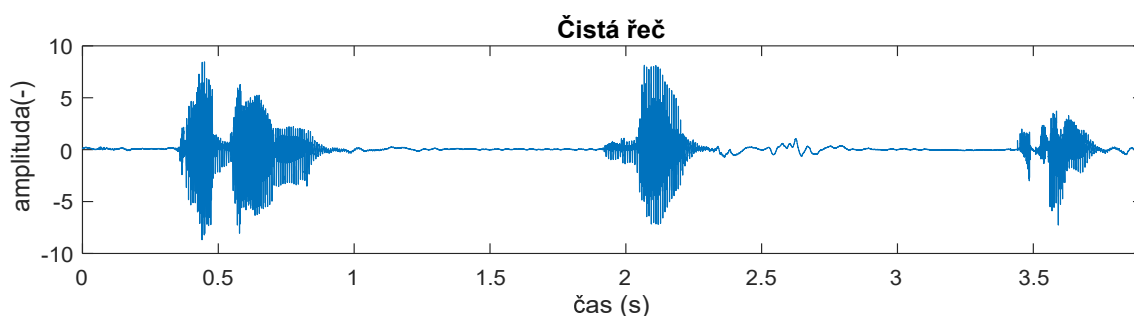
Signály byly systémem zpracovány ve velmi krátkém čase (přibližně tři sekundy). Lze si všimnout, že ANFIS si mnohem lépe poradil se šumem v oblasti vyšších frekvencí, zatímco v oblastech nižších frekvencí tak účinný nebyl. Výstupní hodnota SNR byla 9,54 dB (zlepšení o 18 dB), výstupní hodnota DTW byla 120 (snížení o 1500) a výstupní RMSE bylo 0,24 (snížení o 1,8). Parametr SSNR v tomto testu nebyl určován, jelikož se nejednalo o řečový signál.

Test se sinusovým signálem byl proveden pro demonstrování efektivity ANFIS na různých frekvencích. Z výsledku je vidět dobrý předpoklad pro filtraci řeči, jejíž frekvence se pohybuje ve vyšších oblastech spektra. Podrobnější experimenty se zvukovými nahrávkami budou popsány v následující kapitole.

### 7.3 Filtrace řeči

Cílem provedených experimentů bylo ověření efektivity adaptivního systému pro různé nastavení parametrů filtrace a nalezení optimálního nastavení. Jak bylo uvedeno výše, pro objektivní posouzení efektivity filtrace byly stanoveny hodnoty SNR, SSNR, RMSE a DTW. Na čistě reálných datech však tyto charakteristiky měřit nelze, jelikož v reálných podmínkách nemáme k dispozici čistý řečový signál potřebný pro jejich výpočet. Následující kapitoly se budou věnovat filtraci řeči, pro kterou byl nelineární šum uměle vygenerován.

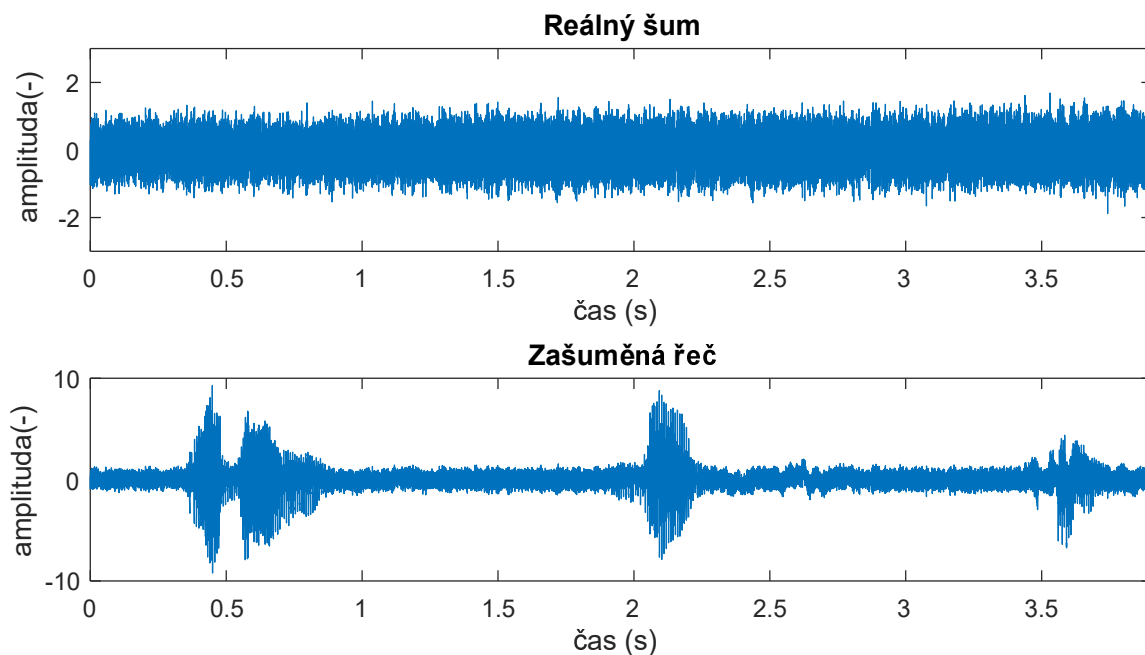
Záznam řeči pro následující experimenty je monofonní a obsahuje 200 000 vzorků. Vzorkovací frekvence signálu je 51 200 Hz, jedná se tedy o záznam trvající necelé čtyři vteřiny. Záznam obsahuje tři slova: "jedna", "dva" a "tři".



Obrázek 23: Čistá řeč

### 7.3.1 Deformace šumu FIR filtrem

V prvním experimentu byl šum nahrán v podobě reálné pořízené nahrávky (zvuk běžícího fénu) se stejnými parametry jako čistá řeč. Zašuměný signál byl v tomto případě vytvořen součtem dvou reálných záznamů a následné transformace FIR filtrem 20. řádu, který tak nasimuloval zkreslení a zpoždění signálů, způsobené průchodem neznámým prostředím. Nutno podotknout, že FIR filtr představuje lineární prvek, který frekvence signálů pouze pozměňuje. Charakter zkreslení signálů je tedy také lineární.



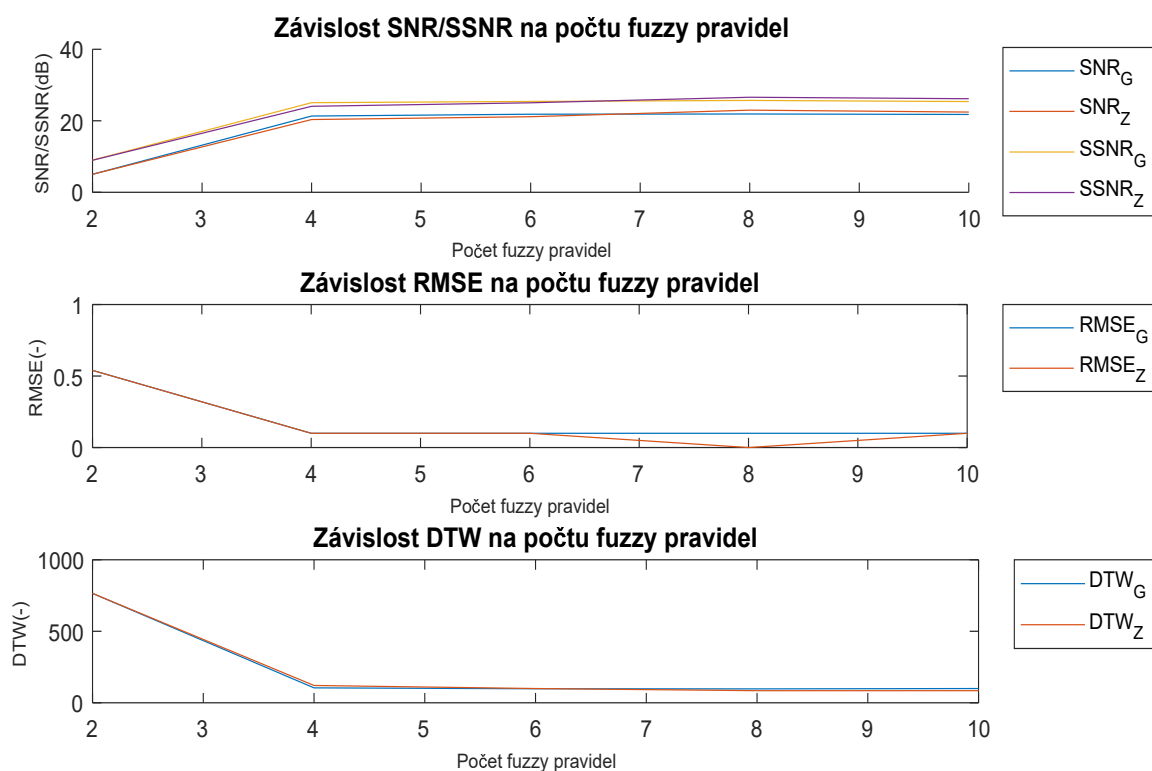
Obrázek 24: Reálný šum (nahore) a zašuměný signál (dole)

Tabulka 3: Vstupní charakteristiky signálu deformované FIR filtrem

SNR (dB)	SSNR (dB)	DTW (-)	RMSE (-)
5,08	8,99	766,3	0,54

V tabulce 3 jsou uvedeny hodnoty SNR, SSNR, DTW a RMSE pro vstupní signál. Šum v tomto případě dosahuje pouze malé amplitudy, což se na těchto hodnotách také odráží.

Provedené experimenty byly zaměřeny na efektivitu filtrace v závislosti na počtu epoch, počtu fuzzy pravidel a typu funkce příslušnosti. Pro tento test byly základní parametry ANFIS nastaveny na dvě epochy, dvě fuzzy pravidla a počet trénovacích vzorků 200 000 (celý signál). Zkoumaný parametr byl poté postupně zvyšován z hodnoty dva na hodnotu deset s krokem dva, zatímco ostatní parametry zůstaly pevné. Test proběhl pro obě implementované funkce příslušnosti, kterými jsou Gaussova a zvonová MF.



Obrázek 25: Závislost na počtu fuzzy pravidel - deformace FIR filtrem

Na výše uvedeném grafu můžeme vidět závislost hodnotících kritérií na počtu fuzzy pravidel. Počet epoch v tomto případě na kvalitu filtrace neměl žádný vliv, proto zde závislost v podobě grafu uvedena není. Opačným případem je počet fuzzy pravidel, který má na kvalitu filtrace velký vliv. Tato závislost je uvedena na obr. 25. Lze si všimnout, že již mezi krokem dva a čtyři dochází k velkému zlomu. Hodnoty SNR a SSNR byly navýšeny o přibližně 16 dB, hodnota RMSE byla snížena o 0,4 a DTW byl snížen o 680. Za tímto zlomem již k výraznějšímu zlepšení nedošlo. Stejný trend je vidět u obou aplikovaných funkcí příslušnosti, mezi kterými byly ve výsledcích minimální rozdíly.

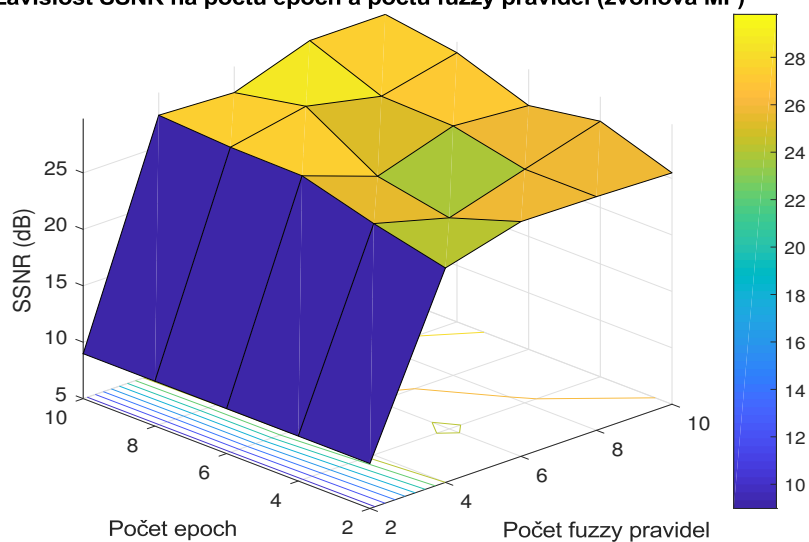
Podrobnější analýzu závislosti na počtu epoch a počtu fuzzy pravidel pro zvonovou MF shrnuje následující tabulka.

Tabulka 4: Výsledky pro jednotlivé kombinace počtu epoch a pravidel (zvonová MF) - deformace FIR filtrem

Počet pravidel	Počet epoch	SNR (dB)	SSNR (dB)	DTW (-)	RMSE (-)
2	2	5,06	8,98	765,17	0,54
2	4	5,06	8,98	765,17	0,54
2	6	5,06	8,98	765,17	0,54
2	8	5,06	8,98	765,17	0,54
2	10	5,06	8,98	765,17	0,54

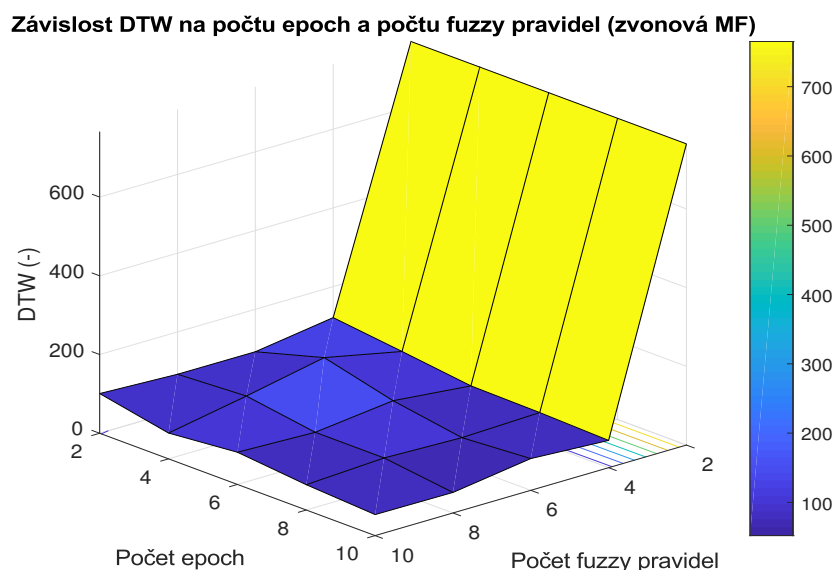
4	2	20,36	24	121,3	0,1
4	4	21,79	25,5	100,3	0,1
4	6	23,49	27,3	77,9	0
4	8	23,5	27,4	75,06	0
4	10	23,9	27,8	69,3	0
6	2	22,09	25,8	92,1	0,1
6	4	26,01	23,7	142,2	0,1
6	6	20,99	24,95	97,7	0,1
6	8	24,44	28,72	76,8	0
6	10	23,5	27,5	80,66	0,1
8	2	21,88	25,7	92,4	0,1
8	4	21,6	25,7	102,06	0,1
8	6	23,26	27,1	74,95	0
8	8	23,4	27,3	74,75	0
8	10	26,22	29,8	52,1	0
10	2	21,82	25,5	92,4	0,1
10	4	23,65	27,64	65,61	0
10	6	23,08	81,8	765,17	0
10	8	24,65	28,85	64,33	0
10	10	26,3	29,81	53,41	0

**Závislost SSNR na počtu epoch a počtu fuzzy pravidel (zvonová MF)**



Obrázek 26: Závislost SSNR na počtu epoch a počtu fuzzy pravidel - deformace FIR filtrem

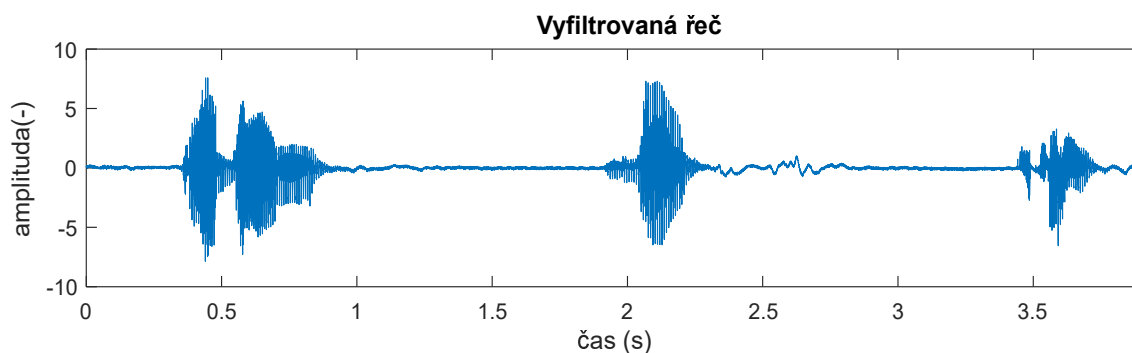




Obrázek 27: Závislost DTW na počtu epoch a počtu fuzzy pravidel - deformace FIR filtrem

Podrobnější analýza závislosti na počtu epoch a počtu fuzzy pravidel naznačuje, že ani pro zvyšující se počet pravidel, počet epoch nemá v tomto případě na celkový výsledek téměř žádný vliv. Rozdíl mezi výsledkem pro nastavení 2/2 a 10/10 je u odstupu signálu od šumu 21,24 dB, u segmentálního SNR 20,83 dB, u DTW 711,76 a u RMSE 0,54.

ANFIS si v tomto experimentu poradil s šumem deformovaným FIR filtrem velmi dobře. Maximálního efektu bylo dosaženo již při základním nastavení ANFIS, přičemž další navyšování parametrů výraznější efekt nemělo. Tohoto výsledku bylo docíleno díky lineárnímu charakteru zkreslení signálu a celkové jednoduchosti dat.



Obrázek 28: Vyfiltrovaná řeč - deformace FIR filtrem

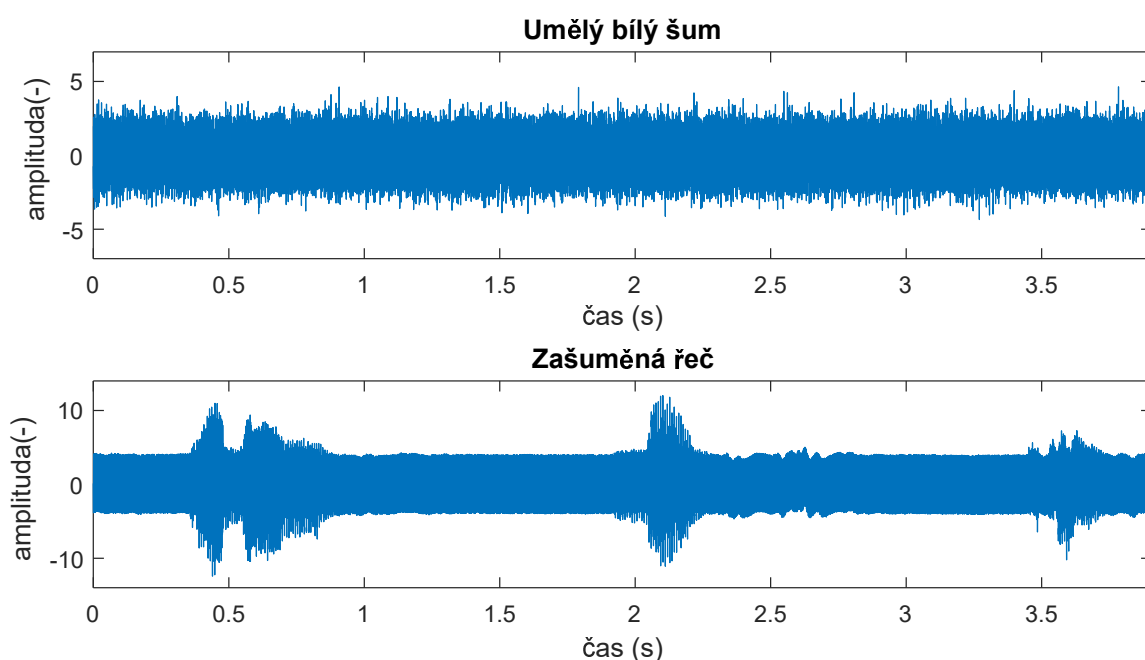
Tabulka 6: Výstupní charakteristiky signálu deformované FIR filtrem

SNR (dB)	SSNR (dB)	DTW (-)	RMSE (-)
26,2	29,81	53,45	0

Na obr. 28 je zobrazen výstupní signál filtrovaný při deseti epochách a deseti pravidlech. Vizuálně můžeme posoudit, že vyfiltrovaná řeč je téměř totožná s původní čistou řečí (viz obr. 23). Výstupní hodnoty jsou uvedeny v tabulce 6.

### 7.3.2 Deformace šumu nelineární funkcí

Pro další experiment na nahrávce řeči byl vytvořen umělý bílý šum, obdobně jako u příkladu se sinusovým signálem. Průchod signálů nelineárním prostředím byl v tomto případě nasimulován transformací bílého šumu nelineární funkcí uvedenou v rovnici 7.6. Tímto způsobem byl vygenerován signál odpovídající šumu po průchodu neznámým prostředím, který byl následně přičten k čisté řeči. V tomto případě byla řeč poškozena šumem mnohem více než v předchozím případě.



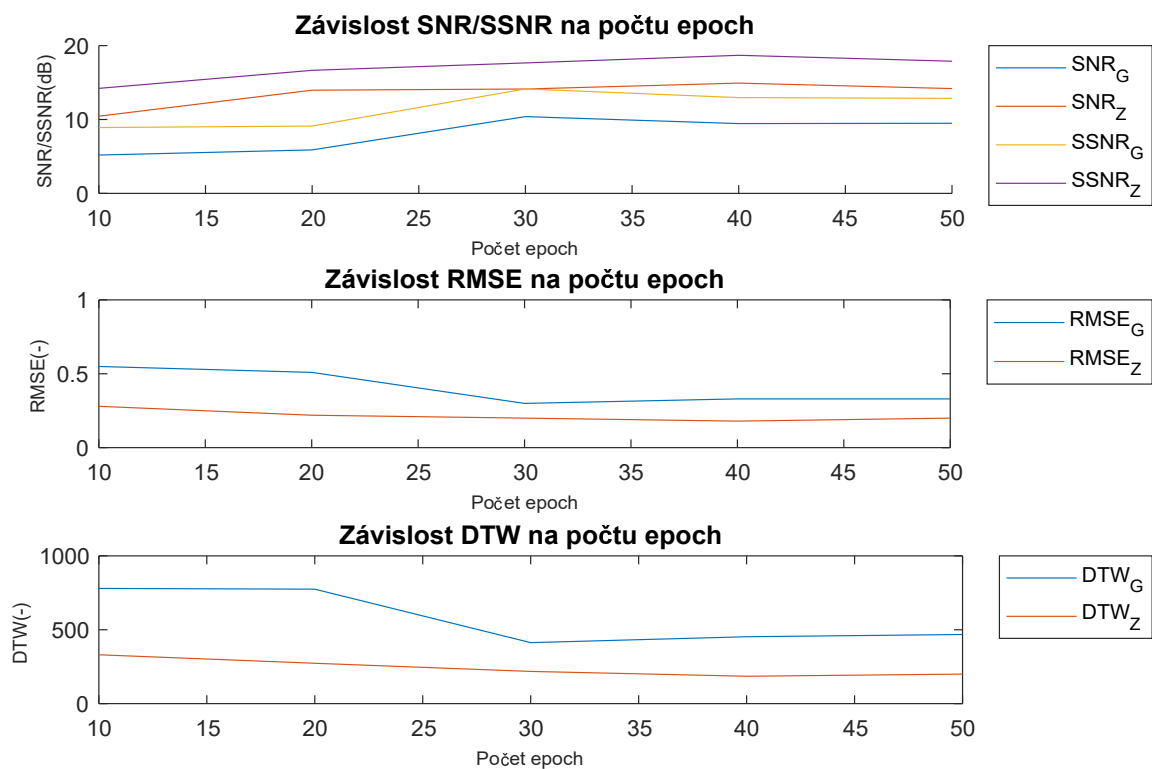
Obrázek 29: Bílý šum (nahore) a zašuměný signál (dole)

Tabulka 7: Vstupní charakteristiky signálu pro vytvořený bílý šum

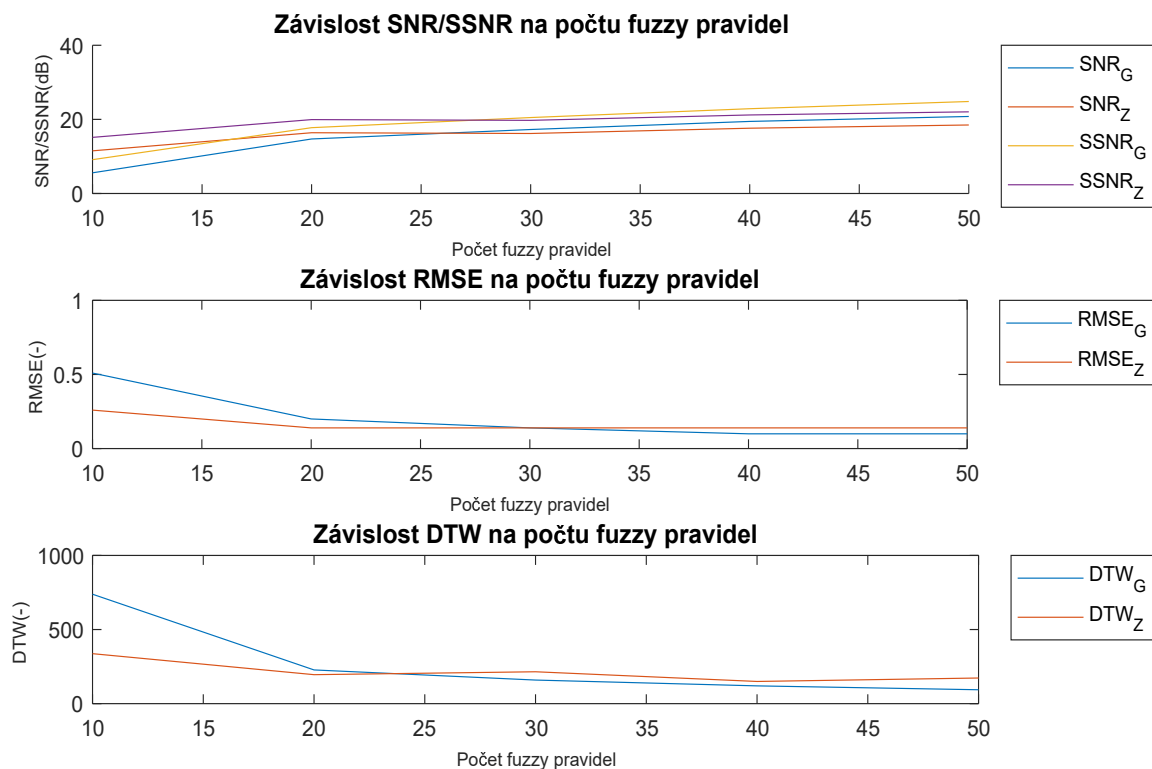
SNR (dB)	SSNR (dB)	DTW (-)	RMSE (-)
-6,64	-3,03	3353,71	2,08

Na obr. 29 lze vidět, že řeč je v šumu tentokrát utopena mnohem více než v prvním experimentu. Sledované parametry byly opět počet epoch, počet fuzzy pravidel a dvě funkce příslušnosti (Gaussova a zvonová). Epochy a pravidla byly tentokrát zkoumány v rozmezí deset až padesát s krokem deset, jelikož šum transformovaný nelineární funkcí představuje pro ANFIS složitější data. Charakter zkreslení je zde nelineární.

Základní nastavení ANFIS v tomto experimentu byl počet epoch deset, počet pravidel deset a počet trénovacích vzorků 200 000 (celý signál).



Obrázek 30: Závislost na počtu epoch - deformace nelineární funkcí



Obrázek 31: Závislost na počtu fuzzy pravidel - deformace nelineární funkcí

Pro složitější data se již projevila závislost kvality filtrace na počtu epoch. Tato závislost je však pouze malá. Rozdíl mezi výstupem pro deset epoch a padesát epoch byl přibližně 3 dB u SNR i SSNR. Projevují se zde už také rozdíly mezi typy MF, kde zvonová funkce dosahuje lepších výsledků než Gaussova funkce. Rozdíl mezi těmito dvěma MF je pro SNR/SSNR přibližně 4 dB.

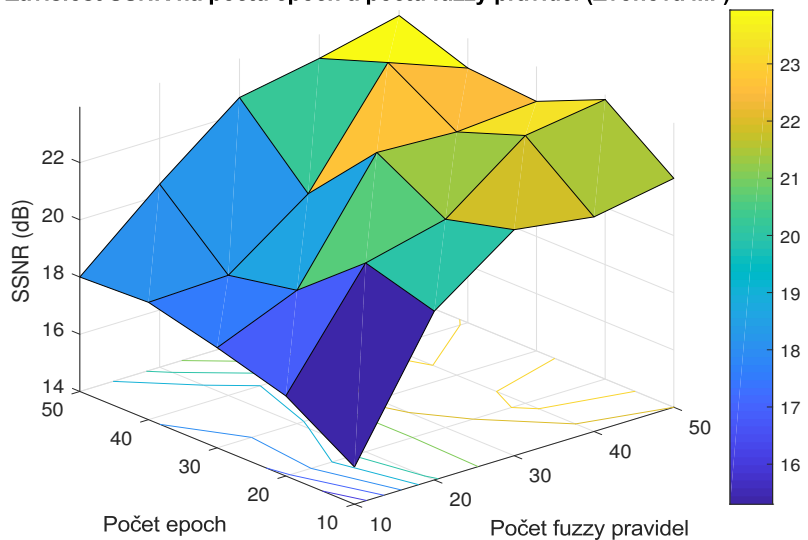
Závislost na počtu fuzzy pravidel byla i v tomto případě velká. Mezi hodnotami deset a padesát dochází k narůstajícímu resp. klesajícímu trendu. Pro Gaussovu MF je rozdíl mezi počáteční a koncovou hodnotou u SNR/SSNR 15 dB, u DTW 644 a pro RMSE 0,41. Pro zvonovou funkci tyto rozdíly nebyly tak markantní. Pro SNR/SSNR byl tento rozdíl 6,5 dB, pro DTW 164 a pro RMSE 0,12.

Tabulka 8: Výsledky pro jednotlivé kombinace počtu epoch a pravidel (zvonová MF) - deformace nelineární funkcí

Počet pravidel	Počet epoch	SNR (dB)	SSNR (dB)	DTW (-)	RMSE (-)
10	10	11,37	15,3	309,2	0,28
10	20	12,9	16,8	282,1	0,22
10	30	14,1	17,5	245,2	0,21
10	40	14,8	18,1	201,42	0,17
10	50	14,6	18	209,7	0,2
20	10	16,8	19,9	208,2	0,17
20	20	18,2	20,6	138,8	0,1
20	30	16,9	18,66	213,25	0,17
20	40	16,1	18,2	252,14	0,2
20	50	16,4	20,4	172,4	0,14
30	10	17,9	21,9	183,1	0,14
30	20	17,74	21,28	158,8	0,14
30	30	19,03	22,63	135,55	0,1
30	40	16,9	20,2	172,4	0,14
30	50	19,25	22,58	121,3	0,1
40	10	19,9	21,2	116,2	0,1
40	20	19,73	23,36	119,8	0,1
40	30	18,82	22,5	135,6	0,1
40	40	20,8	23,94	115,6	0,1
40	50	19,97	23,1	128	0,1
50	10	19	22	162,4	0,1
50	20	20	23,77	120,49	0,1
50	30	19	22,22	140,2	0,1
50	40	19,8	22,9	131,7	0,1

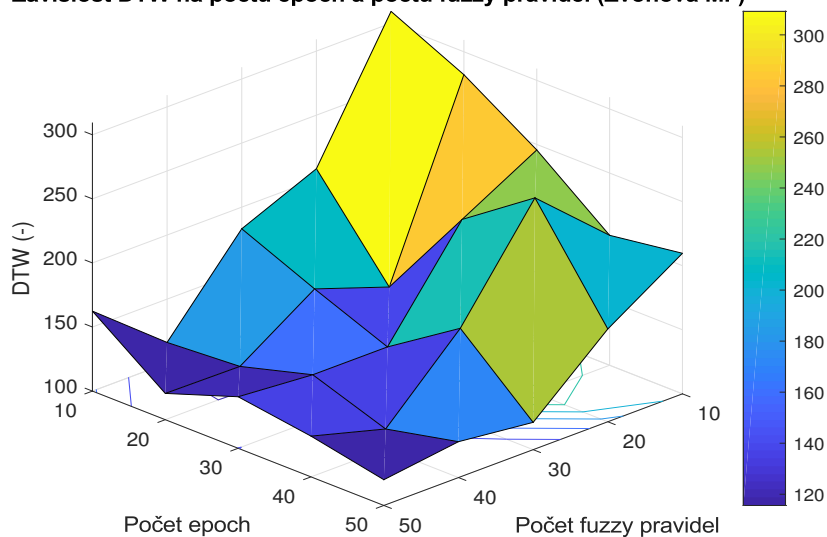
50	50	20,3	23,75	120,1	0,1
----	----	------	-------	-------	-----

**Závislost SSNR na počtu epoch a počtu fuzzy pravidel (Zvonová MF)**



Obrázek 32: Závislost SSNR na počtu epoch a počtu fuzzy pravidel - deformace nelineární funkcí

**Závislost DTW na počtu epoch a počtu fuzzy pravidel (Zvonová MF)**

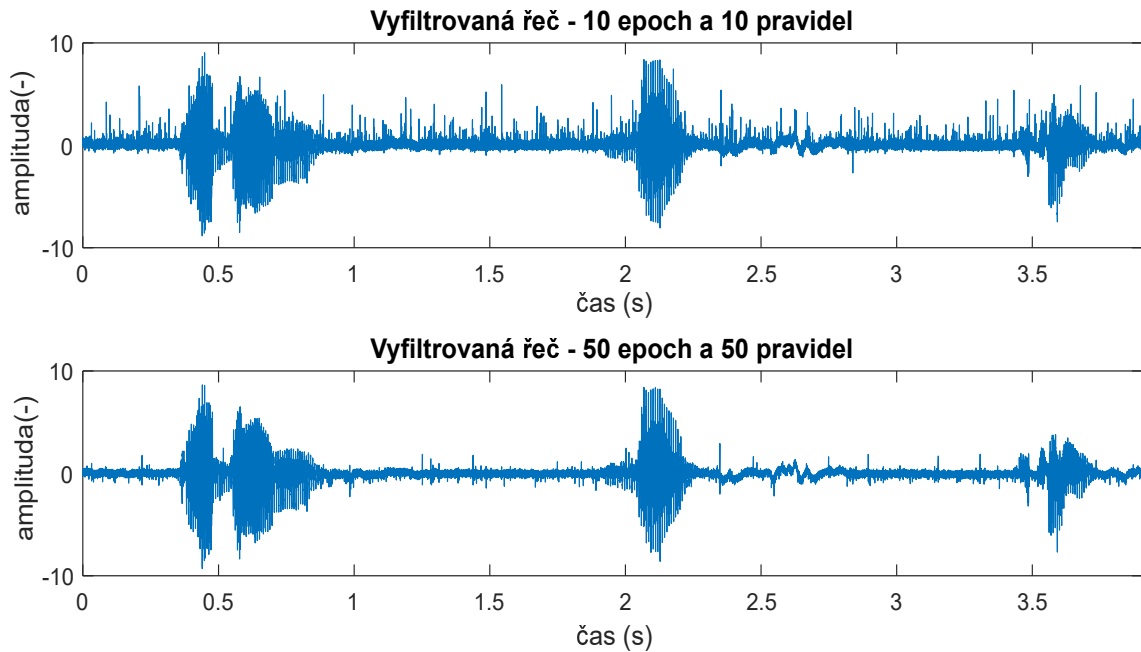


Obrázek 33: Závislost DTW na počtu epoch a počtu fuzzy pravidel - deformace nelineární funkcí

Z uvedených grafů je zřejmé, že při zvyšování počtu epoch i pravidel dochází k lepší filtraci. K tomuto zlepšení dochází však až při vyšších hodnotách parametrů. Pomalejší adaptace ANFIS v tomto experimentu byla způsobena učením struktury na syntetických datech, které byly v testu

použity. Lze si také všimnout, že charakteristika není lineární, ani hladká, ale hodnoty v ní kolísají. Toto kolísání je způsobeno právě horší adaptací ANFIS na umělá data.

Pro nastavení padesát epoch a padesát pravidel bylo zlepšení oproti původním hodnotám 27 dB u SNR/SSNR, 3323 u DTW a 1,98 u RMSE.



Obrázek 34: Vyfiltrovaná řeč - deformace nelineární funkce

Tabulka 10: Výstupní charakteristiky signálů pro vytvořený bílý šum

Počet epoch/počet pravidel	SNR (dB)	SSNR (dB)	DTW (-)	RMSE (-)
10/10	11,77	15,54	310,59	0,28
50/50	18,5	22,15	139,66	0,1

Na obr. 34 jsou zobrazeny výstupní filtrované signály. Pro nastavení deset epoch a deset pravidel je v řeči jasně viditelný šum, který se nepodařilo odfiltrovat. Kvalitnější výsledek byl dosažen pro nastavení padesát epoch a padesát pravidel, kde šum dosahuje velmi malé amplitudy.

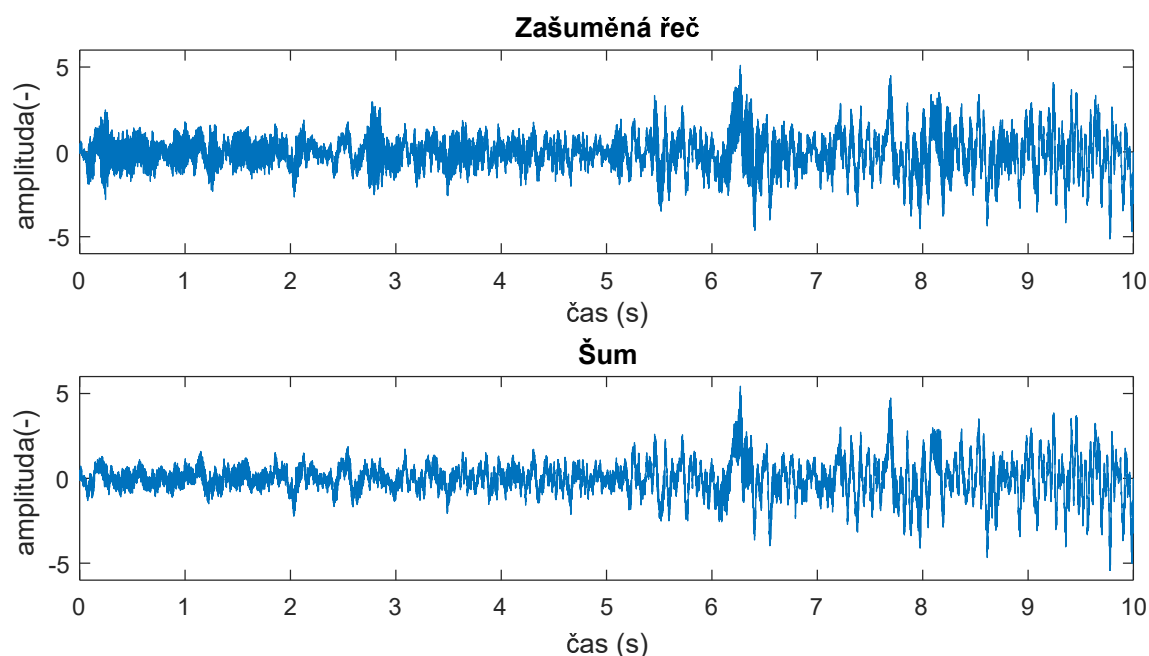
### 7.3.3 Reálná data

Pro experiment s reálnými daty byly využity nahrávky dostupné na webových stránkách [34], obsahující datasety pro evaluaci technik pro odstranění šumu z řeči. Tyto datasety byly vytvořeny smísením čisté řeči, která byla nahrána v tiché kanceláři, a šumu, který představoval ruch náměstí, ruch v kavárně nebo zvuky rozjíždějícího se metra v podzemí. Všechny zvukové záznamy mají vzorkovací frekvenci 16 kHz a jsou stereofonní.

Z těchto datasetů byla vybrána deseti sekundová nahrávka, obsahující čtyři věty čítající celkem třicet slov. Šum zde představuje zvuk rozjíždějícího metra. Záznam má 320 000 vzorků.

Tabulka 11: Vstupní charakteristiky reálných nahrávek

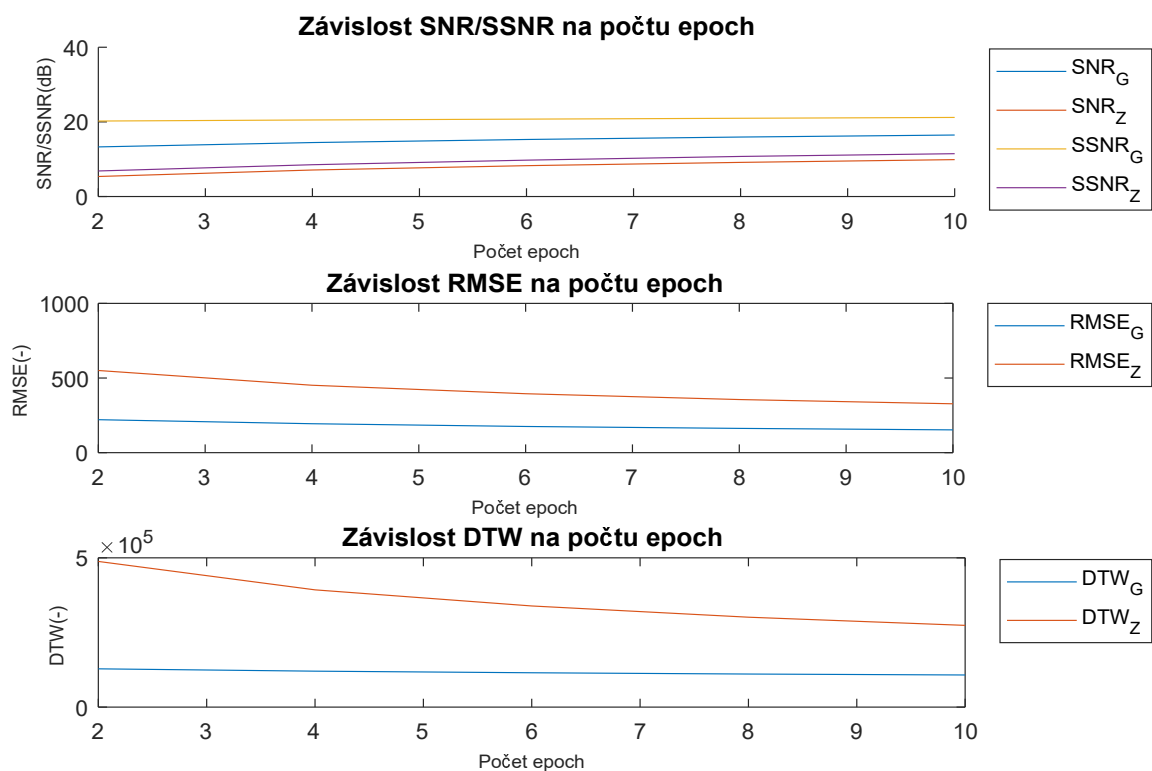
SNR (dB)	SSNR (dB)	DTW (-)	RMSE (-)
-8,88	-5,79	2 931 218	2851,83



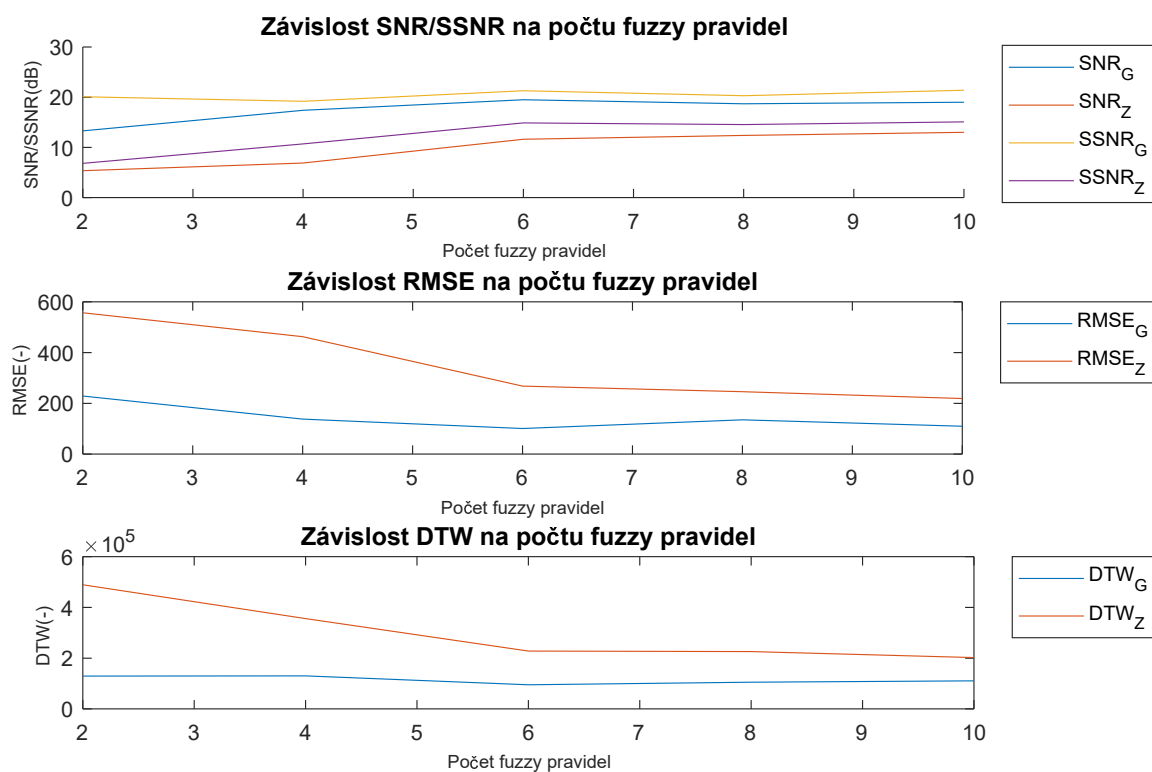
Obrázek 35: Reálné nahrávky

Na obr. 35 jsou zobrazeny časové průběhy vstupních signálů. Šumový signál v průběhu záznamu postupně navyšuje svou amplitudu, čímž dochází na konci nahrávky k většímu poškození řeči. Vlastní řeč je v zašuměném signálu patrná pouze na začátku nahrávky.

Sledovanými parametry byly i v tomto experimentu počet epoch, počet fuzzy pravidel a dvě funkce příslušnosti (Gaussova a zvonová). Počty epoch a pravidel se měnily postupně v rozmezí dva až deset s krokem dva. Nastavení testu bylo tedy stejné jako u experimentu s šumem transformovaným FIR filtrem.



Obrázek 36: Závislost na počtu epoch - reálná data



Obrázek 37: Závislost na počtu fuzzy pravidel - reálná data



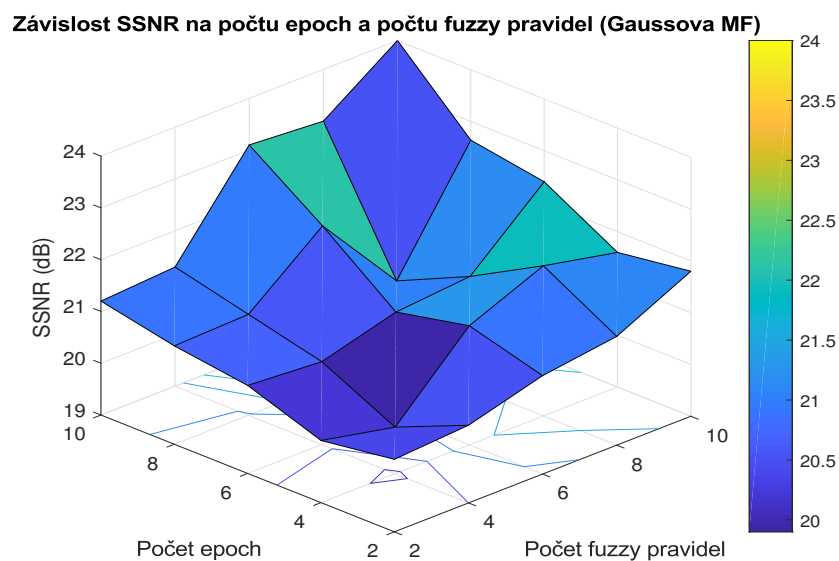
Mezi jednotlivými typy funkcí příslušnosti jsou nyní vidět větší rozdíly než tomu bylo u předchozích dat. Gaussova MF dosáhla oproti zvonové MF lepších výsledků u všech hodnotících kritérií. Pro odstup signálu od šumu byl tento rozdíl přibližně 8 dB v základním nastavení dvě epochy a dvě pravidla.

Počet epoch měl zde větší vliv na kvalitu filtrace, a to zejména u zvonové MF. U Gaussovy funkce bylo zlepšení pouze mírné, přičemž u odstupu signálu od šumu bylo potlačení šumu soustředěno především na oblasti neobsahující řeč (SSNR bylo téměř konstantní).

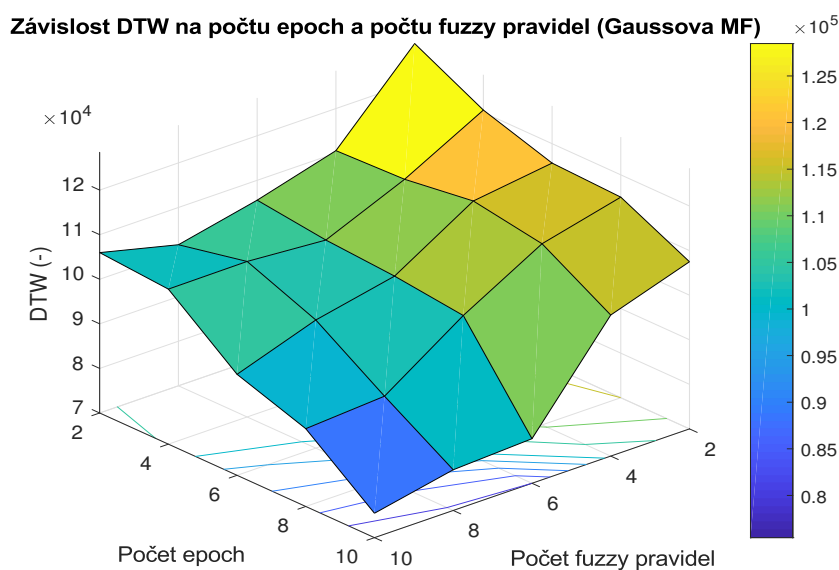
U závislosti na počtu pravidel je patrný postupný nárůst resp. pokles až do hodnoty kroku šest. Od této hodnoty již další navyšování počtu pravidel výraznější vliv nemělo.

Tabulka 12: Výsledky pro jednotlivé kombinace počtu epoch a pravidel (Gaussova MF) - reálná data

Počet pravidel	Počet epoch	SNR (dB)	SSNR (dB)	DTW (-)	RMSE (-)
2	2	13,3	20,4	128 474	221,2
2	4	14	20,2	120 548	215,12
2	6	15,3	20,7	115 143	176,5
2	8	15,5	20,9	114 999	168,6
2	10	16,5	21,2	107 449	153,97
4	2	15,31	20,5	110 555	215,54
4	4	17,3	19,9	111 914	120,55
4	6	16,9	20,6	113 291	150
4	8	17,11	20,95	110 645	170
4	10	18,36	21,3	101 644	102,78
6	2	18,5	20,9	105 543	203,7
6	4	17,97	21,3	103 588	185,9
6	6	18,9	21	102 344	175,01
6	8	19,54	22,1	100 645	143,3
6	10	21,9	23,1	79 912	70,5
8	2	19	21,1	101 645	157,6
8	4	17,9	21,9	104 888	143,17
8	6	19,67	21,13	98 745	100,66
8	8	20	20,48	88 641	91
8	10	21,5	23	79 125	71,9
10	2	19	21,4	115 977	130,1
10	4	21,3	21,6	104 777	123,1
10	6	21,3	22,4	92 655	94,05
10	8	21,4	22,64	87 454	85
10	10	22,5	24	75 412	67,2

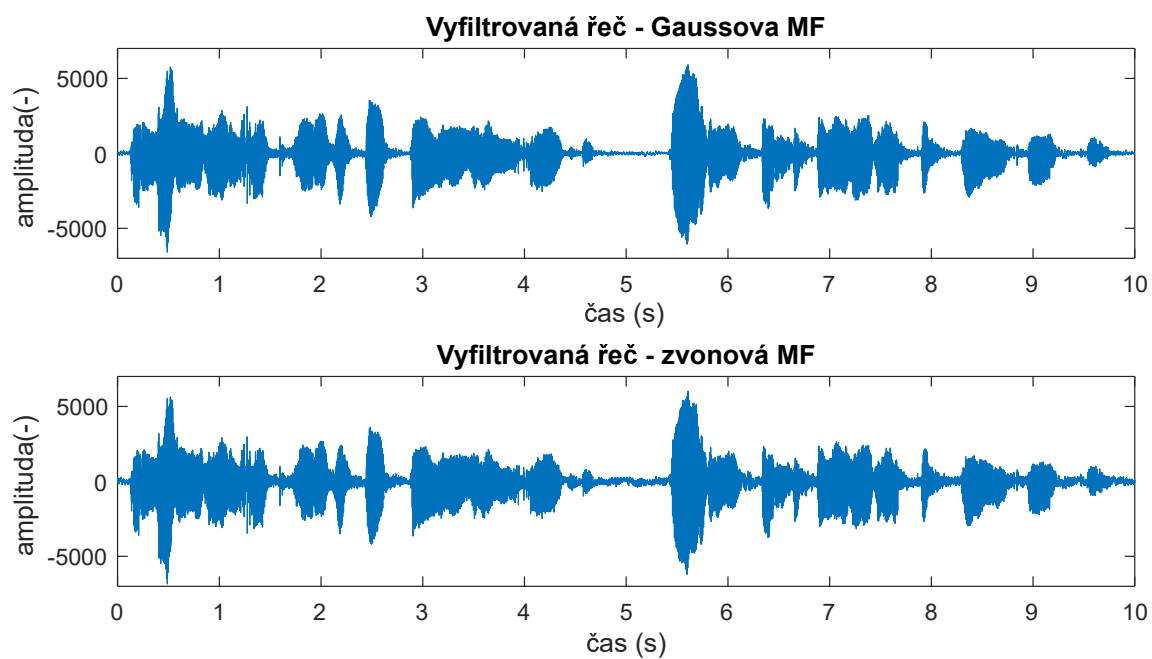


Obrázek 38: Závislost SSNR na počtu epoch a počtu fuzzy pravidel - reálná data



Obrázek 39: Závislost DTW na počtu epoch a počtu fuzzy pravidel - reálná data

Kvalita filtrace je v tomto případě výrazně ovlivněna oběma parametry. Mírnou převahu mají fuzzy pravidla, jejichž vliv se zde projevuje zejména při počtu deseti epoch. Celkové zlepšení mezi minimálním a maximálním nastavením ANFIS je zde 9,2 dB u SNR, 3,6 dB u SSNR, 53 026 u DTW a 154 pro RMSE.



Obrázek 40: Vyfiltrovaná řeč - Gaussova a zvonová MF

Na grafickém zobrazení vyfiltrovaných signálů je vidět, že zejména v oblastech neobsahující řeč, je u zvonové MF patrný malý šum, který se s touto funkcí nepodařilo odfiltrout.

Tabulka 14: Výstupní charakteristiky signálů pro reálná data

Typ MF	SNR (dB)	SSNR (dB)	DTW (-)	RMSE (-)
Gaussova	22,5	24	75 412	67
zvonová	14,5	16,2	192 564	225

## 8 Zhodnocení výsledků

V této kapitole budou shrnuty získané poznatky z provedených experimentů. Tyto experimenty byly zaměřeny na zjištění závislosti zkoumaných parametrů s ohledem na výslednou kvalitu filtrace a realizovatelnost systému. Zkoumanými parametry byly počet epoch, počet fuzzy pravidel a typ funkce příslušnosti, kterými byly Gaussova a zvonová MF. V rámci experimentů byly otestovány celkem tři typy dat:

- Data deformovaná FIR filtrem,
- data deformovaná nelineární funkcí,
- reálná data.

V prvním testu vstupní data představovaly reálně pořízené nahrávky. Zašuměný signál byl následně vytvořen součtem těchto signálů a transformací přes FIR filtr 20. řádu.

Test s těmito daty prokázal, že již při nastavení dvě epochy a čtyři fuzzy pravidla, byl efekt filtrace na maximální možné míře. Další navyšování parametrů nebo změna typu MF na výslednou kvalitu filtrace již vliv prakticky neměla. ANFIS se na tuto deformaci signálu dokázal přizpůsobit velmi rychle. Trénovací sadou zde byla reálná nahrávka šumu na pozadí, která v kombinaci s lineárně deformovaným signálem přinesla maximální možný výsledek ve velmi krátkém čase.

V druhém experimentu byl šumový signál uměle vygenerován a poté transformován nelineární funkcí. Zde se již vliv parametrů na výslednou kvalitu filtrace prokázal. ANFIS se na tyto data adaptoval hůře, než tomu bylo v ostatních experimentech, což je přisuzováno syntetické povaze dat, na kterých byl ANFIS učen.

Zkoumané parametry byly postupně navyšovány z hodnoty deset na hodnotu padesát. Z naměřených závislostí lze vyvodit, že kvalitu filtrace ovlivňují všechny parametry. Počet epoch zde měl jen malý vliv, v rozmezí deset až padesát doznala kvalita filtrace jen mírného zlepšení. Zvonová funkce příslušnosti dosáhla oproti Gaussové funkci lepších výsledků. Největší vliv se však prokázal u fuzzy pravidel, přičemž platí, že se zvyšujícím se počtem pravidel dochází ke zlepšování filtrace. Na druhou stranu je nutné podotknout, že počet pravidel má také velmi negativní vliv na výpočetní náročnost systému. V níže uvedené tabulce jsou vypsány časy zpracování signálů o celkové délce 200 000 vzorků.

Tabulka 15: Čas zpracování

Počet epoch	počet pravidel	čas (s)
10	10	21
10	50	263
50	10	76
50	50	1080

Vzhledem k realizovatelnosti systému je vhodné zvolení nižšího počtu epoch a fuzzy pravidel. Optimální nastavení by v tomto případě představovalo použití třiceti epoch a dvaceti pravidel, což jsou hraniční hodnoty, od kterých další navyšování výraznější zlepšení už nepřineslo.

Poslední experiment byl zaměřen na filtraci reálných dat. Vliv jednotlivých parametrů zde byl obdobný jako u předchozího případu. Lepších výsledků zde dosáhla Gaussova MF. Z hlediska kvality filtrace bylo uspokojivého výsledku dosaženo již při nižším nastavení hodnot parametrů, jelikož trénovací sadou zde byla reálná nahrávka šumu, na kterou se ANFIS dokázal lépe přizpůsobit. Hodnoty epoch a fuzzy pravidel byly zkoumány v rozmezí dva až deset. Optimálním nastavením pro reálná data tak představovalo šest fuzzy pravidel a libovolný počet epoch při použití Gaussovy funkce příslušnosti. Nutno podotknout, že toto optimum je zvoleno pro použitá reálná data, a pro jiné nahrávky může být odlišné. To je dáno zejména charakterem prostředí, ve kterém je nahrávka pořízena.

## Závěr

Cílem této diplomové práce bylo vytvoření softwarové realizace systému ANFIS v programovacím jazyce C# a jeho využití v oblasti zpracování řečové nahrávky.

Teoretická část této práce byla zaměřena na rozbor problematiky zpracování řeči, rozbor metod adaptivní filtrace a samotného systému ANFIS. Ze znalostí získaných v této části byl následně navržen adaptivní systém, který byl poté implementován v programovacím jazyce C#. Vytvořená aplikace umožňuje uživateli provést offline filtraci zvukových nahrávek a export vyfiltrovaného signálu. Přínosnou funkcionalitou aplikace je možnost nalezení optimální hodnoty požadovaného hodnotícího parametru na zvoleném parametru filtru.

Efektivita vytvořené aplikace byla ověřena na různých typech dat, zahrnující syntetická data i reálná data. Provedené experimenty kladly důraz na výslednou kvalitu filtrace a výpočetní náročnost pod různým nastavením parametrů ANFIS. Zkoumanými parametry byly počet epoch, počet fuzzy pravidel a typ funkce příslušnosti. Pro objektivní posouzení efektivity byly v každém testu stanoveny hodnotící kritéria SNR, SSNR, DTW a RMSE.

Po zhodnocení získaných výsledků lze říci, že navržený adaptivní systém dosahuje uspokojivé kvality filtrace. Je však nutné brát v potaz charakter zpracovávaných dat a správné nastavení parametrů ANFIS, které mohou mít jak pozitivní, tak negativní vliv na efektivitu filtrace. Tento negativní vliv se projevuje zejména v celkové výpočetní náročnosti.

Přínosem této diplomové práce je důkladné zhodnocení závislosti parametrů ANFIS na výslednou kvalitu filtrace pro různý charakter zkoumaných signálů. V problematice potlačení šumu z řečového signálu je tato práce dobrým základem pro další výzkum a využití adaptivního neuro-fuzzy inferenčního systému, a to zejména v reálných aplikacích s možností rozšíření systému o schopnost zpracování signálů v reálném čase.

## Literatura

- [1] JANG, J.-SR. ANFIS: adaptive-network-based fuzzy inference system. *IEEE transactions on systems, man, and cybernetics*, 1993, 23.3: 665-685.
- [2] MARTINEK, Radek, et al. Adaptive noise suppression in voice communication using a neuro-fuzzy inference system. In: *Telecommunications and Signal Processing (TSP)*, 2015 38th International Conference on. IEEE, 2015. p. 382-386.
- [3] KUMAR, Jay, et al. Environmental Noise Cancellation by Using Neuro Fuzzy Adaptive Filtering. In: *Communication Systems and Network Technologies (CSNT)*, 2015 Fifth International Conference on. IEEE, 2015. p. 1157-1162.
- [4] PANDEY, Bipul, et al. Multilingual speaker recognition using ANFIS. In: *Signal Processing Systems (ICSPS)*, 2010 2nd International Conference on. IEEE, 2010. p. V3-714-V3-718.
- [5] LU, Yang; HAITAO, Yang; HUAIRONG, Shen. A comparative analysis on HHT and ANFIS in nonlinear noise cancellation. 2010.
- [6] MIRY, Mohammed Hussein; MIRY, Ali Hussein; KHLEAF, Hussain Kareem. Adaptive noise cancellation for speech employing fuzzy and neural network. In: *Energy, Power and Control (EPC-IQ)*, 2010 1st International Conference on. IEEE, 2010. p. 289-296.
- [7] YANG, Lu; YANG, Haitao; SHEN, Huairong. The Application of ANFIS and WT in Filtering. In: *Information Engineering and Computer Science (ICIECS)*, 2010 2nd International Conference on. IEEE, 2010. p. 1-3.
- [8] THEVARIL, Jasmin; KWAN, H. K. Speech enhancement using adaptive neuro-fuzzy filtering. In: *Intelligent Signal Processing and Communication Systems*, 2005. *ISPACS 2005. Proceedings of 2005 International Symposium on*. IEEE, 2005. p. 753-756.
- [9] Encyklopedie fyziky. Fyzika :: MEF [online]. [cit. 2017-11-29]. Dostupné z: <http://fyzika.jreichl.com>
- [10] Kinematika [online]. [cit. 2017-11-29]. Dostupné z: [http://ufmi.ft.utb.cz/texty/env\\_fyzika/EF\\_03.pdf](http://ufmi.ft.utb.cz/texty/env_fyzika/EF_03.pdf)
- [11] VLACHÝ, Václav. *Praxe zvukové techniky*. 2. aktualiz. vyd. Praha: Muzikus, c2000. ISBN 80-86253-05-8.
- [12] Signal-to-noise ratio. Scholarpedia [online]. [cit. 2017-12-02]. Dostupné z: [http://www.scholarpedia.org/article/Signal-to-noise\\_ratio](http://www.scholarpedia.org/article/Signal-to-noise_ratio)
- [13] Tutorial (Fuzzy Logic Toolbox) [online]. [cit. 2017-12-05]. Dostupné z: <https://edoras.sdsu.edu/doc/matlab/toolbox/fuzzy/fuzzytut.html>

- [14] POKORNÝ, M., SROVNAL, V. Systémy s umělou inteligencí - Učební text a návody do cvičení. CZ.1.07/2.2.00/15.0113. VŠB - Technická univerzita Ostrava. Ostrava, 2012
- [15] PENHAKER, Marek a Martin AUGUSTYNEK. Zdravotnické elektrické přístroje 1. Ostrava: Vysoká škola báňská - Technická univerzita Ostrava, 2013, 411 s. ISBN 978-80-248-3107-7.
- [16] RABINER, Lawrence R.; SCHAFER, Ronald W. Theory and applications of digital speech processing. Upper Saddle River: Pearson, 2011.
- [17] MARTÍNEK, Radek. Využití adaptivních algoritmů LMS a RLS v oblasti adaptivního potlačování šumu a rušení. 2013.
- [18] ŠKUTOVÁ, Hana. Extrakce plodového elektrokardiogramu metodami soft computingu [online]. Ostrava, 2016 [cit. 2017-12-25]. Diplomová práce. VŠB - Technická Univerzita Ostrava. Vedoucí práce Ing. Radek Martínek, Ph. D.
- [19] MOLDŘÍKOVÁ, Zuzana. Metody potlačení šumu pro rozpoznávače řeči [online]. Brno, 2014 [cit. 2017-12-25]. Diplomová práce. Vysoké učení technické v Brně. Vedoucí práce Ing. Jan Odstrčilík.
- [20] MARTINEK, Radek, et al. A robust approach for acoustic noise suppression in speech using ANFIS. Journal of Electrical Engineering, 2015, 66.6: 301-310.
- [21] Adaptivní filtrace zašuměných řečových signálů. Elektrorevue [online]. Brno [cit. 2017-12-26]. Dostupné z: <http://www.elektrorevue.cz/clanky/02063/index.html>
- [22] MARTINEK, Radek; ZIDEK, Jan. The real implementation of NLMS channel equalizer into the system of software defined radio. Advances in Electrical and Electronic Engineering, 2012, 10.5: 330.
- [23] ZELINKA, Ivan. Umělá inteligence, aneb, Úvod do neuronových sítí, evolučních algoritmů. Vyd. 2. Zlín, 2005. ISBN ISBN 80-7318-277-7.
- [24] VONDRÁK, Ivo. Umělá inteligence a neuronové sítě. Ostrava, 2009. ISBN 80-7078-259-5.
- [25] The McCulloch-Pitts Model of Neuron. Department of Electrical and Computer Engineering [online]. Texas, 2008 [cit. 2017-12-27]. Dostupné z: <http://www.ece.utep.edu/research/webfuzzy/docs/kk-thesis/kk-thesis-html/node12.html>
- [26] MARTINEK, Radek. The Use of complex adaptive methods of signal processing for refining the diagnostic quality of the abdominal fetal cardiogram. Ostrava, 2014. Disertační práce. Vysoká škola báňská - Technická univerzita Ostrava. Vedoucí práce Jan Žídek.
- [27] ZADEH, Lotfi A. Fuzzy sets. Information and control, 1965, 8.3: 338-353.



- [28] MARTINEK, Radek, et al. Non-invasive fetal monitoring: A maternal surface ECG electrode placement-based novel approach for optimization of adaptive filter control parameters using the LMS and RLS algorithms. *Sensors*, 2017, 17.5: 1154.
- [29] VONDRASEK, M. a P. POLLAK. Methods for Speech SNR Estimation: Evaluation Tool and Analysis of VAD Dependency. *Radioengineering* [online]. Společnost pro radioelektronické inženýrství, 2005, 14(1), 6-11 [cit. 2018-03-04]. ISSN 1210-2512. Dostupné z: <http://hdl.handle.net/11012/57991>
- [30] JANG, Jyh-Shing Roger; SUN, Chuen-Tsai; MIZUTANI, Eiji. Neuro-fuzzy and soft computing; a computational approach to learning and machine intelligence. 1997.
- [31] UPADHYAY, Navneet; KARMAKAR, Abhijit. Speech enhancement using spectral subtraction-type algorithms: A comparison and simulation Study. *Procedia Computer Science*, 2015, 54: 574-584.
- [32] SOUNDARYA, S.V.S a Rajani ALUGONDA. A Review on Various Speech Enhancement Techniques. *International Journal of Advanced Research in Computer and Communication Engineering*. 2016, 5(8), 1-6. ISSN 2278-1021.
- [33] HARDING, Philip John. Model-Based Speech Enhancement. Norwich, 2013. Disertační práce. University of East Anglia. Vedoucí práce Dr. Ben Milner.
- [34] Two-channel mixtures of speech and real-world background noise. *SiSEC 2011* [online]. [cit. 2018-04-15]. Dostupné z: <http://sisec2011.wiki.irisa.fr/tiki-index79f3.html?page=Two-channel+mixtures+of+speech+and+real-world+background+noise>

## A Obsah přiloženého CD

- DP\_HOL0267\_Holub.pdf – PDF dokument obsahující text diplomové práce
- ANFIS\_HOL0267\_Holub – adresář se zdrojovými kódy aplikace